PHYSICAL MAPPING AND CHARACTERIZATION OF THE EQUINE LYMPHOCYTE ANTIGEN (ELA) COMPLEX

A Dissertation

by

ASHLEY GUSTAFSON SEABURY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2005

Major Subject: Genetics

PHYSICAL MAPPING AND CHARACTERIZATION OF THE EQUINE LYMPHOCYTE ANTIGEN (ELA) COMPLEX

A Dissertation

by

ASHLEY GUSTAFSON SEABURY

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Loren Skow (Chair of Committee)	Bhanu Chowdhary (Member)
James Womack (Member)	Ian Tizard (Member)
Geoffrey Kapler (Chair of Genetics Faculty)	Evelyn Tiffany-Castiglioni (Head of Department)

May 2005

Major Subject: Genetics

ABSTRACT

Physical Mapping and Characterization of the Equine Lymphocyte Antigen (ELA) Complex. (May 2005)

Ashley Gustafson Seabury, B.S., West Virginia University

Chair of Advisory Committee: Dr. Loren C. Skow

The major histocompatibility complex (MHC) is a genomic region comprised of a linked cluster of genes and gene families that play an important role in both the adaptive and innate immune responses. Genes within the MHC have also been associated with susceptibility and/or resistance to certain diseases, such as haemochromatosis, insulin-dependent diabetes, and psoriasis. As a result of these associations the MHC is one the most extensively studied regions of the mammalian genome. The MHCs of a wide variety of species, such as human (HLA), mouse (H-2), pig (SLA), and cow (BoLA), have been characterized with respect to gene content, genomic organization of class I, class II, and class III regions, and comparative organization. Comparative analyses have been useful in delineating the evolutionary development of the MHC.

While the MHC of many mammalian species has been investigated, little research has been performed on the equine (*Equus caballus*) MHC. The equine MHC is referred to as the equine lymphocyte antigen (ELA) complex and is located on chromosome ECA20q. The research that has been done on ELA focused on identifying gene copy number and genetic polymorphisms in the classical class I and class II genes. To better

characterize the gene content and organization of ELA, we isolated 103 bacterial artificial chromosome (BAC) clones from a horse BAC library containing well conserved genes found within mammalian MHCs. These BAC clones were assembled into two sequence-ready ordered contigs that span the ELA complex. The first contig which has a minimum tiling path of nine BAC clones contains the ELA class II region and spans 800 kb. The class I and III regions are contained within the second contig which has a 14 BAC clone minimum tiling path and spans 1.6 Mb. This study will report on the construction of the two BAC contigs which span the ELA complex, and characterization of the gene content and organization of the ELA complex.

DEDICATION

This dissertation is the culmination of countless hours of research, none of which would have been possible without my mother, husband, and family. My mother gave me the courage and confidence to strike out on my own and pursue my dreams. My husband Chris has given me four years of support, encouragement, and reassurance. Thank you Mom and Chris!

ACKNOWLEDGEMENTS

I am grateful to my committee chair, Dr. Loren Skow for his guidance and support. I thank Dr. Bhanu Chowdhary for having confidence in my abilities and for constantly encouraging me. Furthermore, I thank Dr. James Womack and Dr. Ian Tizard for serving on my graduate committee.

Many people have helped me in the laboratory over the past four years. I am forever grateful to Dr. Terje Raudsepp for performing all the fluorescent in situ hybridization (FISH) experiments. I would like to thank my fellow graduate student Nicole Ramlachan for all her assistance both in and out of the laboratory over the past three years. I want to acknowledge and thank my husband, Christopher Seabury, for all of his support and for developing an efficient BAC end sequencing strategy. Finally, a special thanks to our former lab manager, Dee Honeycutt, and our student workers Daniel McGilvray and Emily Zayac who have helped out in the lab in any way possible.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

History of the MHC

The major histocompatibility complex (MHC) is a genomic region comprised of linked clusters of genes and gene families and is one of the most extensively researched regions of the mammalian genome. The MHC was first identified in the early 1900's through a series of tumor grafting experiments in mice (Little and Tyzzer 1916). Results of allogenic tumor grafting experiments (Little and Tyzzer 1916) and subsequent studies of skin transplantation in identical mice (Bover 1927) strongly suggested that several genes determined the transplant outcome (acceptance or rejection of the transplanted tissue). In the late 1930's Peter Gorer identified an antigen that strongly effected transplantation outcomes. This antigen was called antigen-II and marks the first identification of the mouse MHC (Gorer 1937). While Gorer investigated the mouse MHC from an immunological perspective, George Snell searched for the genetic factors controlling transplant rejection through genetics. Since it was difficult to differentiate between the many H (histocompatibility) loci, Snell decided the only way to study a single H locus was to develop strains (congenic) of mice that were identical at all loci, except the H locus of interest (Klein 2001). Eventually, Snell developed a congenic strain of mice whose H locus co-segregated with the *Fused* mutation, thus allowing Snell to identify

---

This dissertation follows the style and format of the journal Immunogenetics.

a unique H locus, which was eventually termed H-2 (Gorer et al. 1948; Snell 1948).

The human MHC, called the human leukocyte antigen (HLA) system, was first identified in the 1950's by identification of HLA antigens based upon observations of agglutination of leukocytes (Dausset 1958; Payne and Rolfs 1958; van Rood et al. 1958). Initial studies of HLA were based upon HLA serological typing which defined antigens on leucocytes based upon agglutination studies. However this process of HLA sero-typing proved to be difficult due to the high degree of polymorphism in the HLA genes. Despite the difficulties associated with sero-typing, it was the primary method of analysis for HLA until the advent of modern molecular techniques such as polymerase chain reaction (PCR), restriction fragment length polymorphism (RFLP), and DNA sequencing. These molecular techniques have allowed HLA and other vertebrate MHCs to be studied with respect to genomic organization and evolution, frequency of polymorphism, and associations with disease.

Genetic organization of the MHC

The MHC is a genomic region that contains a collection of linked genes and gene families, many of which play an important role in the innate and adaptive immune response. Moreover, the MHC is one of the most gene dense regions in the mammalian genome. The HLA region contains one gene every 18 kb and 40% of the expressed genes in this region are functionally involved in the immune response (Kulski et al. 2002). The genetic organization of the MHC has been analyzed and characterized in several species ranging from the nurse shark (Flajnik et al. 1999) to humans (MHC

Sequencing Consortium 1999). It is through these comparative studies that the conserved organizational features of the MHC have been defined. One of the primary organizational features of the MHC is its division into three non-overlapping regions (class I, class II, and class III). The class I and class II regions were defined based upon the functional characteristics of the genes contained within them, while the class III region was initially defined as the region intercalated between them. Additionally, further analysis of the regions adjacent to the MHC indicated that there is a considerable degree of conserved synteny, high linkage disequilibrium, distinct differences in the isochore structure that exist within these regions (Herberg et al. 1998; Malfroy et al. 1997; Stephens et al. 1999; Totaro et al. 1998a; Totaro et al. 1998b; Yoshino et al. 1997; Zemmour et al. 1990) Taken together this information was utilized to expand the traditional MHC. The extended class II region was in part designated based upon the identification dramatically different isochores found within the class II region. For example in humans, the traditional class II region was found to be located within a low GC content L2 isochore, while the adjacent region, which was named the extended class II region, was determined to be within a high GC content H2 isochore (Stephens et al. 1999). The extended class I and class II regions were identified through a variety of different means and their exact boundaries are not clear and vary from publication to publication. Thus, for the purpose of this text the extended MHC regions will be defined as they are in the review by Horton et al. (2004) . This information can also be found on the Sanger web site for the extended MHC (http://www.sanger.ac.uk/HGP/Chr6/XMHC/index.shtml). The extended class I spans

from *MOG* to *SCGN* and the extended class II region extends from *COL11A2* to *SYNGAP1* (Horton et al. 2004).

Class I region

The class I region was initially defined as the region containing class I genes. The class I genes encode cell surface glycoproteins that form heterodimers with non-MHC endcoded ß2-microglobulin protein. The class I heterodimer presents peptides derived from intracellular proteins to CD8+ cytotoxic T-cells, thus playing a key role in the adaptive immune response. Class I molecules also interact with natural killer (NK) cells to prevent inappropriate NK cell driven cell lysis (Reyburn et al. 1997). Upon further investigation of the function, expression, and polymorphisms of the class I genes, these genes were further divided into class Ia, Ib, and Ic genes. The class Ia genes, which are referred to as the classical class I genes, are expressed on all nucleated cells and demonstrate a high degree of polymorphism in the regions encoding the antigen binding site (ABS). These are the class I genes that present intracellularly derived antigenic peptides to CD8+ cytotoxic T-cells (Hughes et al. 1999; Klein and O'hUigin 1994). The class Ib genes are referred to as the non-classical class I genes and are characterized as being much less polymorphic than the class Ia loci and are expressed in a tissue-specific manner (Hughes et al. 1999; Klein and O'hUigin 1994). The specific function of the class Ib genes has yet to be elucidated. The class Ic genes also have relatively little polymorphism relative to the class Ia genes and appear to have diverged before the radiation of placental orders (Hughes et al. 1999). This class designation is less

commonly used than the class Ia and Ib, nonetheless it does describe yet another group of class I genes found within this region. Additionally, class I pseudogenes and gene fragments exist within the class I region. These pseudogenes and gene fragments are non-functional; however they likely play a role in the generation of diversity in the functional class I genes through interlocus gene conversion (Grimsley et al. 1998).

The number of classical and non-classical class I genes as well as class I pseudogenes is highly variable across species. For instance, humans have six functional class I genes (Flajnik et al. 1999), while the mouse has approximately 30 functional class I genes (Amadou et al. 1999; Kumanovics et al. 2002). This variability among species is likely the result of the manner by which the class I genes have evolved. Class I genes have evolved through a series of insertion, duplication, deletion, and reorganizing events that ultimately resulted in the accumulation of multiple functional and non-functional class I genes (Hughes and Nei 1988; Hughes and Yeager 1998b; Yeager and Hughes 1999). The precise series of events that occurred during the molecular evolution of the class I genes in one particular species is not likely to be shared among divergent species. This is illustrated by the inability to determine orthology between class I genes from divergent species (Hughes and Nei 1989a; Hughes and Yeager 1998b; Yeager and Hughes 1999).

A large number of non-class I genes are also found in the class I region of most species. The non-class I genes have a variety of functions varying from transcription factors (*POU5F1* and G*TFIIH*) to cell growth proteins (*TCF19*). Ironically, the non-class I genes outnumber the class I genes in this region that was defined by their

presence. However, it appears that the non-class I genes play an important role in determining the genetic organization of the class I region. The significance of the non-class I genes was elucidated through comparative analyses of the class I region in multiple species. These analyses revealed that the non-class I gene are orthologous across different mammalian species and generally existed in the same relative order (Amadou 1999; Amadou et al. 1999). The primary perturbation of orthology in the class I region is the result of differences in the number and location of the class I genes (Kulski et al. 2002). Comparative analyses of the human and mouse class I regions could not determine orthology with respect to the class I genes. However, the class I genes were found to exist in homologous regions within the class I region (Amadou 1999). Amadou (1999) developed the "framework hypothesis', which proposes that the non-class I genes represent a framework or scaffold for the class I region, to explain these observations. A series of "permissive" sites (or duplication blocks) exist within this framework that allows the insertion and subsequent duplication and expansion of class I genes (Amadou 1999). These "permissive" sites are defined by the genes which flank them and have been found to be conserved across different species (Amadou 1999). There are three primary "permissive" sites in the class I region that have been designated duplication block alpha (*TCTEX4* to *MOG*), beta (*BAT*1 and *POU5F1*), and kappa (*GNL1* to *TRIM26*) (Amadou 1999; Kulski et al. 2001; Kulski et al. 2002; Kumanovics et al. 2003).

Class II region

Like the class I region, the class II region was characterized by the presence of functional class II genes, with class II genes encoding both alpha and beta cell surface glycoproteins. Functional class II molecules form a heterodimer from the alpha and beta class II molecules. The class II heterodimers, which are expressed on antigen presenting cells (macrophages, dendritic cells, and B-cells), present antigenic peptides derived from phagocytosed foreign extracellular antigens to CD4+ T cells (Kumanovics et al. 2003). The class II molecules are synthesized in the endoplasmic reticulum and targeted by the invariant chain to endocytic pathway. The invariant chain occupies the peptide binding groove of the class II molecule and stabilizes the folded protein until the antigenic peptide can be loaded with the help of the non-classical class II gene *DM* (Busch et al. 2000; Kumanovics et al. 2003; Watts 2001). Additionally, the non-classical class II gene *DO* modulates the loading of antigenic peptides by the *DM* molecules onto class II molecules in B cells (Alfonso and Karlsson 2000; Brocke et al. 2002).

Five orthologous groups of class II genes (*DP*, *DM*, *DO*, *DQ*, *DR*) are conserved among mammals (Kumanovics et al. 2003). The *DP*, *DQ*, and *DR* loci are referred to as the classical class II genes and encode receptors that present antigenic peptides to CD4+ T cells. The *DO* and *DM* loci are non-classical class II loci and are involved in peptide loading of the class II molecules (Kulski et al. 2002). The class II molecules have arisen through a process of duplication and subsequent divergence, resulting in extensive genetic variation in the number of loci, both functional and pseudogenes (Yeager and Hughes 1999). Interestingly, the *DM* and *DO* genes do not appear to have pseudogenes,

unlike the *DP*, *DQ*, and *DR* class II genes which possess one to three pseudogenes per gene family in mammals (Kumanovics et al. 2003).

A number of the non-class II genes within the class II region encode proteins that are involved in class I antigen processing and transport. The *PSMB8* and *PSMB9* genes encode proteins that are components of the interferon-γ-inducible proteasome, which generates antigenic peptides presented by the class I molecules (Goldberg and Rock 1992; Monaco 1992a; Monaco 1992c), while the *TAP1* and *TAP2* genes encode proteins that transport antigenic peptides from the cytoplasm to the endoplasmic reticulum for assembly on class I molecules (Monaco 1992b; Neefjes et al. 1993; Shepherd et al. 1993). The genomic location of these class I antigen processing and transport genes is variable and species-dependent. In the mammalian MHC these genes are tightly linked to the class II region, however in the avian and teleost MHCs these genes are linked to the class I region. Additionally, there are non-class II genes such as death associated protein 6 (*DAXX)*, collagen type XI, alpha 2 (*COL11A2*), and butyrophilin-like 2 (*BTNL2*) that play no role in the immune response, yet are highly conserved and maintained within the class II region of the vertebrate MHC (Beck et al. 2001).

Class III region

The class III region is the smallest and most gene dense region of the MHC, which is located interstitial to the class I and class II regions. Additionally, the class III region is the most highly conserved region of the MHC. In humans the class III region contains one gene every 14.5 kb (Shiina et al. 2004a). The *NOTCH4* (Notch homolog 4) gene is

the first class III gene that marks the boundary between the class II and class III regions, while *BAT1* (HLA-B associated transcript 1) marks the boundary between the class III and class I regions (Xie et al. 2003). The class III region contains a variety of genes with different functions including but not limited to roles in the immune response, in particular the complement system and the inflammatory response. For example, *TNF* (tumor necrosis factor alpha), *LTA* (lymphotoxin alpha), *C2* (complement component 2), *C4* (complement component 4), and *BF* (complement factor B) are class III genes, which are involved in the inflammatory and/or the complement portions of the immune response (Klein and Sato 1998).

The class I and class II regions are characterized by gene families that have undergone repeated duplication and divergence to produce multiple copies of functional genes and nonfunctional pseudogenes, but the class III region is almost completely devoid of duplicated genes and pseudogenes. The only gene duplication identified in the class III region is the *C4* - *CYP21A2* (steroid 21-hydroxylase) segment. Comparative analyses have identified a *C4* - *CYP21A2* segmental duplication in human, mouse, and rat (Blanchong et al. 2001; Kumanovics et al. 2003; Walter et al. 2002; Yang and Yu 2000). For example, in rat an additional copy of the C4, CYP21A2, and STK19 genes were found in the class II region, while in mouse the C4-Cyp21A2 segment is duplicated in tandem (Walter et al. 2002). Interestingly, the specific units which are duplicated vary and the duplication event appears to be species specific, but it appears that all duplication events center upon the *C4* gene (Blanchong et al. 2001). While the *C4* - *CYP21A2* segment is found in many species, the number of *C4* – *CYP21A2* duplicates

often varies depending upon haplotype (Collier et al. 1989; Dunham et al. 1989; Kumanovics et al. 2003; Partanen et al. 1989). The variation in the number of copies of the *C4* gene within species may be the result of unequal meiotic crossover(s) within this region (Kawaguchi et al. 1992; Kumanovics et al. 2003; Sinnott et al. 1990; Tusie-Luna and White 1995). No duplication of the *C4* - *CYP21A2* segment has been identified in the pig (Peelman et al. 1996) and the pig MHC is characterized by a near absence of meiotic recombination (Chardon et al. 2000).

MHC and disease

Initial studies of the MHC in humans and mice were designed to characterize genetic factors that controlled rejection of tumor or tissue transplants, but they quickly identified associations of disease susceptibility or resistance with genes found within the MHC. In humans, MHC-encoding genes have been associated with more than 100 diseases including arthritis, diabetes, haemochromatosis, psoriasis, schizophrenia, cancers, lupus (SLE), coeliac disease, multiple sclerosis, and other autoimmune disorders (Choonhakarn et al. 2002; Dawkins et al. 1999; Dawkins et al. 1983; McCann et al. 1984; Tay et al. 1997), making the MHC a region of considerable biomedical interest. While the MHC region has been associated with many different disease, most associations of disease susceptibility or resistance have focused on haplotypes defined by class I and class II loci. Some good examples of disease resistance and/or susceptibility associations with single HLA alleles occur in humans. For example, haplotypes defined by HLA-B*27 and *57 alleles are associated with a slowed

progression of HIV infection to AIDS (Migueles et al. 2000). However, the unequivocal identification of specific causal genetic variation resulting in disease susceptibility is difficult due to the high levels of polymorphism, segmental duplication, linkage disequilibrium, and frequency of paralogous genes which exist (Apanius et al. 1997).

Interestingly, not all MHC-related disease associations are linked to class I and class II defined haplotypes. In humans the class III region has been associated with predisposed disease susceptibility in some autoimmune diseases, such as lupus erythematosis (Gruen and Weissman 2001; Yu and Whitacre 2004), rheumatoid arthritis (Okamoto et al. 2003) and type 1 diabetes (Nishimura et al. 2003). In fact some genes that are directly related to disease susceptibility have been identified. For example, mutations in the *CYP21A2* gene are associated with congenital adrenal hyperplasia (Mornet et al. 1991; Tusie-Luna and White 1995), while mutations in the *TNX* gene are associated with one form of Ehlers-Danos Syndrome (Burch et al. 1997).

Disease association studies are most easily studied in the human MHC, however MHC associations with disease have been observed in other mammalian species. For example, specific *BoLA-DRB3* alleles were determined to be associated with resistance to persistent lymphocytosis that is caused by the bovine leukemia virus (BLV) (Xu et al. 1993). Interestingly, the strength of the resistance is influenced by the haplotype in which the alleles are found (Lewin et al. 1999; Xu et al. 1993). Additionally, susceptibility to mastitis in cattle was associated with haplotypes that have one set of class II *DQ* genes (Park et al. 2004). Another example of a MHC disease association in mammals exists in horses. The predisposition for the development of sarcoid tumors

was associated with certain MHC serological haplotypes (Brostrom et al. 1988; Chambers et al. 2003; Goodrich et al. 1998; Meredith et al. 1986). In particular the A3W13 serological haplotype existed more frequently in horses that developed sarcoid tumors (Brostrom et al. 1988), nonetheless an association with a specific MHC gene, allele, or haplotypes has yet to be identified. However, association studies performed in vertebrate species are complicated by high levels of polymorphism, linkage disequilibrium, and the presence of multiple genes that are members of a gene family, as is the case in human disease association studies.

Comparative analyses of the MHC

Extensive gene maps and sequences have been developed for the MHCs of several species due to a variety of associations such as disease resistance and/or susceptibility, mate selection, embryonic development, and evolution. Gene maps and genomic sequence are available for regions of the MHC in a variety of species such as human (MHC Sequencing Consortium 1999), mouse (Amadou et al. 1999; Kumanovics et al. 2002; Xie et al. 2003), cat (Beck et al. 2001; Yuhki et al. 2003), rat (Gunther and Walter 2001; Hurt et al. 2004; Ioannidu et al. 2001; Walter et al. 2002), pig (Chardon et al. 2001; Renard et al. 2001), chicken (Kaufman et al. 1999b), quail (Shiina et al. 1999a; Shiina et al. 2004b), zebrafish (Kuroda et al. 2002), and medaka (Nonaka et al. 2001). The availability of gene maps and genomic sequence for a variety of species that have diverged at different times permits comparative studies of the organization, gene content, gene order, and evolution of the MHC region. (Flajnik and Kasahara 2001; Flajnik et al.

1999; Kasahara et al. 2004). These analyses help predict how the MHC region has changed and evolved since its origins ~ 520 million years ago as well as the likely ancestral organization of this region. Interestingly, comparative analysis of the MHC multiple species revealed that the MHC is found in all jawed vertebrates examined, but not in jawless vertebrates or invertebrates (Flajnik and Kasahara 2001). Thus, it appears that the MHC and the adaptive immune system emerged approximately 520 million years ago when the first jawed vertebrate appeared (Flajnik 1998; Flajnik and Kasahara 2001; Flajnik et al. 1999).

The human MHC region, which is referred to as the human leukocyte antigen (HLA), was one of the first MHC regions to be discovered and characterized. Thus, it has served as a framework or "gold standard" for the analysis and characterization of the MHC region in other species. HLA is located on the short arm of human chromosome 6 (HSA 6p21.3) and spans 3.6 Mb (MHC Sequencing Consortium 1999). HLA contains around 239 genes including 130, 17 gene candidates, 4 non-coding genes and 88 pseudogenes (Shiina et al. 2004a). Approximately 40% of the expressed genes encode proteins that function in the innate and/or adaptive immune system (Trowsdale 2001). Additionally, HLA has a large number of duplicated and highly polymorphic genes, for example the class I A locus has over 250 alleles (Forbes and Trowsdale 1999).

HLA is divided into the class I, class II, and class III regions. Each region contains a diverse array of genes and unique characteristics. Additional research has also expanded the boundaries of the traditional MHC, resulting in the development of the extended class I and extended class II regions (Horton et al. 2004).

The class I region contains ~18 class I genes of which six are coding of which 12 are pseudogenes, as well as seven MHC class I –chain-related (MIC) genes (MHC Sequencing Consortium 1999; Shiina et al. 2004a; Shiina et al. 1999b). These genes are distributed within the alpha, beta, and kappa duplication blocks of the class I region (Kulski et al. 2002). A variety of non-class I genes are found within the class I region whose functions relate to the cell growth (tripartite motif-containing 10 gene,*TRIM10*), DNA replication and repair (mediator of DNA damage checkpoint 1 gene, MDC1), transcriptional regulation (POU domain, class 5, transcription factor 1 gene, *POU5F1*), and chaperone function and signaling (flotillin 1 gene, *FLOT1*) (Shiina et al. 2004a; Shiina et al. 1999b; Shiina et al. 1999c). These genes have been identified in the same general order in most mammalian MHCs studied to date, which supports the framework hypothesis discussed previously. The extended class I region expands the traditional class I region to the *SCGN* gene and contains a variety of genes including class I-like genes, such as the hemochromatosis gene (*HFE*), histone genes, and olfactory receptor genes (Horton et al. 2004; Totaro et al. 1998b). Interestingly, the largest histone gene cluster in the human genome is located in the extended class I region (Horton et al. 2004; Marzluff et al. 2002)

The class II region of HLA bounded by the *BTNL2* gene and *DP* genes, contains 19 class II genes of which six are pseudogenes (Shiina et al. 2004a). The class IIb α/β gene pairs *DO* and *DM* have no related pseudogenes, while the class IIa α/β gene pairs *DP*, *DQ*, and *DR* have multiple pseudogenes (MHC Sequencing Consortium 1999; Shiina et al. 2004a). The *BRD2* gene is the only gene in the class II region that is not

associated with the immune response (Shiina et al 2004a). Despite the lack of immune function the *BRD2* gene is highly conserved from mammals to jawed fish (Kulski et al. 2002). The extended class II region, which is defined as existing between *COL11A2* and *SYNGAP1*, contains 21 genes of which 15 are expressed (Horton et al. 2004; Shiina et al. 2004a).

The HLA class III region is the most gene dense region of the human genome containing approximately one gene per 14.5 kb (MHC Sequencing Consortium 1999; Shiina et al. 2004a). The class III region spans ~700 kb is located between the class I and class II regions. This region contains 62 genes some of which are involved in the innate immune response, such as members of complement cascade and members of the inflammatory response (Shiina et al. 2004a). Other genes in the class II region are involved in a variety of non-immune related functions. Comparative analysis of the human, mouse, rat, and pig class III region revealed a high degree of conservation of gene content, gene order, and gene architecture indicating that the class III region has remained relatively stable in the genomes of these species since their last common ancestor (Peelman et al. 1996; Walter et al. 2002; Xie et al. 2003). This degree of conservation is particularly remarkable given the variability that is observed in the class I and class II regions. However, additional comparative analyses of the class III regions of both mammalian and non-mammalian MHCs have revealed the existence of some genes that are only found in the mammalian MHC. This observation has been explained by predicting that some of the mammalian MHC specific genes found within the class III

region are non-functional (Mallya et al. 2002; Xie et al. 2003) or are novel and rapidly evolving (Xie et al. 2003).

Size of the MHC region

Comparisons with respect to the size of the MHC region are primarily focused on mammals; however some size estimates do exist for some non-mammalian vertebrates. The human and rodent MHCs, which span approximately 3.5- 4.0 Mb, appear to be the largest of the mammalian species investigated to date (Amadou et al. 1999; MHC Sequencing Consortium 1999; Gunther and Walter 2001). The smallest mammalian MHC identified is in the pig MHC (SLA) that spans approximately 2.4 Mb of DNA, excluding the centromeric region, which disrupts the contiguous MHC (Chardon et al. 1999a; Smith et al. 1995). However, the chicken MHC (B locus) spans 92 kb and is the smallest MHC characterized in any vertebrate species (Kaufman et al. 1999b). Some of the variation in size is accounted for by differences in the number of genes that exist within the MHC regions. This idea is supported by the fact that class I and class II genes within the MHC have undergone a series of duplication, divergence, and/or deletion events that result in the expansion and/or contraction of the MHC region (Yeager and Hughes 1999). The B locus contains only 19 genes and is proposed to represent the minimal essential MHC where all genes dispensable to a functional antigen presentation system were lost (Kaufman et al. 1999b). This is in stark contrast to HLA which contain 239 genes (Shiina et al. 2004a).

Gross genomic organization

The genomic organization of the class I, class II, and class III regions is generally contiguous in the following order: class I / class III / class II. This organization is considered to be the standard conserved gross genomic organization of the mammalian MHC, which is observed in humans and species of Canidae (MHC Sequencing Consortium 1999; Wagner et al. 1999). However, deviations from this organization do exist both as disruptions of the conserved linkage and as the existence of additional class I or class II genes in other areas of the genome.

Disruptions of the contiguous organization of the MHC exist in some mammalian species. In the pig, SLA is found on swine chromosome 7 (Sscr 7) and it is disrupted by the centromere. The class II region is found on Sscr 7q1.1, while the class I and class III regions are found on Sscr 7p1.1 (Smith et al. 1995). The disruption of the MHC by the centromere occurred at the class II – class III boundary resulting in all class II genes existing on the q arm of swine chromosome 7 (Chardon et al. 1999b). Another mammalian species whose MHC organization is disrupted is the cow. The gross genomic organization of the bovine MHC is dramatically different than most other mammals. The bovine MHC (BoLA), which is located on bovine chromosome 23, is disrupted as a result of a chromosomal inversion that translocated a portion of the class II region near the centromere (Andersson et al. 1988; Band et al. 1998). This disruption is different from the one that occurred in pigs because it actually separates genes within a single region as opposed to just separating two regions from each other. Additional investigation through FISH analysis demonstrated that the inversion is present in all

ruminants examined to date, and that the organization of the cattle MHC is likely an ancestral condition for all advanced pecorans (McShane et al. 2001).

The mouse and rat MHCs, H-2 and RT1 respectively, are organized in the conserved class I / class III / class II organization. However, they also have additional class I genes in the extended class II region and olfactory receptor gene regions (Kulski et al. 2002; Kumanovics et al. 2003). The class I gene cluster that exist proximal to the class II and class III regions is referred to as H2-K in mouse and RT1-A in the rat. These regions exist at orthologous regions on the chromosome and these regions are flanked by the same genes (Walter and Gunther 2000). However, these class I regions are not orthologs with respect to their gene content (Hurt et al. 2004).

Further examples of deviations from the conserved organization for the MHC can be observed when comparative analyses are performed with non-mammalian vertebrate species. The MHC has been extensively examined in the chicken (*Gallus domesticus*). The chicken has two MHC regions, the B locus and the Rfp-Y, which are separated by a nuclear organizer region (NOR) on microchromosome 16 (Miller et al. 1996). Within the B locus the class I and class II regions are not clearly defined and only one class III loci (*C4*) that has been identified (Kaufman et al. 1999a; Kaufman et al. 1999b). Analysis of the quail (*Coturnix japonica*) MHC reveals that it has an organization similar to that of the B locus, yet with a dramatic expansion of class I and class II genes (Shiina et al. 2004b). Thus, the genomic organization of the avian MHC is different from that of its mammalian counterparts, however it still maintains the close linkage of the class I and class II loci.

The organization and characteristics of the MHC has also been investigated in a number of different species from widely separated orders of Teleosti, such as such as zebrafish (Bingulac-Popovic et al. 1997; Graser et al. 1998; Kuroda et al. 2002), stickleback, common guppy, cichlid fish (Sato et al. 2000), rainbow trout (Phillips et al. 2003), pufferfish (Clark et al. 2001; Sambrook et al. 2002), and medaka (Nonaka et al. 2001). In all teleost species investigated to date the class I and class II genes of the MHC have been found in separate linkage groups. This organization of the MHC is dramatically different from that found in tetrapods (amphibians, birds, and mammals) where the class I and class II regions are maintained in close linkage. Additional studies of the MHC in teleost have revealed that the class III genes (*BF* and *C4* genes were investigated) do exist in teleost species, however they are not linked to the class I or class II genes (Kuroda et al. 2000; Nonaka et al. 2001; Sambrook et al. 2002).

The MHC has not been found in any jawless fish or invertebrate studied to date and it appears that the MHC and the adaptive immune system arose in a common ancestor of all jawed vertebrates. These conclusions were supported by the identification and characterization of the MHC in different species of cartilaginous fish, which are members of the oldest taxon of extant jawed vertebrates (Ohta et al. 2002; Ohta et al. 2000). Studies of nurse shark and banded houndsharks revealed close linkage of class II α and β genes with each other as well as with the classical class I gene (Ohta et al. 2000). These results indicate that the class I and class II loci have remained linked on a contiguous segment of a chromosome in the cartilaginous fish. Additional experiments on the banded hound shark determined that the class III genes *C4* and *BF* do

exist in tight linkage with the class I and class II regions (Terado et al. 2003). These results confirm that the organization of cartilaginous fish MHC is similar to that observed in tetrapods. Comparative analysis of cartilaginous fish, teleost, and tetrapods indicates that linkage of the class I and class II genes likely represents the ancestral organization of the MHC (Flajnik and Kasahara 2001; Kulski et al. 2002; Ohta et al. 2000). Comparative analysis suggests that the conservation of class I and class II linkage is the result of a primordial duplication in cis (Ohta et al. 2000).

Comparative genetic content

Class I region

Initial comparative analyses of the human and mouse class I region revealed that class I genes were not orthologous, yet they exist in homologous positions within the class I region across mammalian species (Amadou 1999; Amadou et al. 1999). Further comparative studies revealed that the non-class I genes were orthologous across different mammalian species and were located in a relatively conserved order (Amadou 1999; Amadou et al. 1999; Kulski et al. 2002). The "framework hypothesis" explains the conservation of areas within the class I region remain recognizable despite class I gene expansion (Amadou 1999). The "framework hypothesis" proposes that the non-class I genes represent an organizational framework for the class I region and a set of flexible sites, referred to as duplication blocks, exist within this framework to allow the insertion, expansion, and/or deletion of class I genes (Amadou 1999). Three primary duplication blocks: alpha (*TCTEX4* to *MOG*), beta (*BAT*1 and *POU5F1*), and kappa (*GNL1* to

TRIM26) are defined by the non-class I genes that flank them and are conserved among different mammalian species (Kulski et al. 2000; Kulski et al. 2001; Kulski et al. 2002). However, additional class I genes have been identified in the extended class II region and in the olfactory receptor gene regions of the MHC in rodents (Gunther and Walter 2001; Kulski et al. 2002; Kumanovics et al. 2002; Kumanovics et al. 2003). The extended class II region has been investigated in several species and the insertion of class I genes into this region has only be observed in the mouse and rat. These results suggest that the class I insertion into this region may have occurred in a common ancestor of the mouse and rat (Stephens et al. 1999; Walter and Gunther 2000).

Variation of this duplication block organization has been described for the pig MHC. Comparative analysis of the pig class I region revealed that the alpha duplication block is devoid of any class I genes in SLA (Chardon et al. 2001), while the beta and kappa duplication blocks do contain class I genes (Renard et al. 2001; Velten et al. 1999). Thus, all of the pig class I genes exist within beta and kappa duplication blocks in SLA (Chardon et al. 2001). Additionally, the pig beta and kappa duplication blocks and their surrounding non-class I genes are conserved within the HLA class I region, thus indicating that the only difference between the pig and other mammalian class I regions is that SLA has not utilized the alpha duplication block (Chardon et al. 2001; Renard et al. 2001; Velten et al. 1999).

Although, the position of the class I duplication blocks is conserved across mammalian species, the number of class I loci that exist within each block is highly variable. For example in the kappa block HLA has four class I genes, SLA has seven

class I genes, and H-2 and RT1 have more than 50 class I genes (Amadou et al. 1999; Renard et al. 2001; Shiina et al. 1999b; Shiina et al. 1999c). Comparative studies of the class I duplication blocks indicate that the rodent MHC appears to have undergone the most extensive class I gene expansion of all mammalian species studied to date (Kulski et al. 2000; Kulski et al. 2002). However, the exact number of class I genes found within each duplication block varies dependent upon haplotype in both humans and rodents (Kumanovics et al. 2003).

Variation in gene content within the relatively conserved non-class I framework includes the presence of MHC class I –chain-related (*MIC*) genes, which are a type of class Ib gene. The *MIC* genes are found in the class I region of all primates as well as pigs (Shiina et al. 1999b; Shiina et al. 1999c; Velten et al. 1999). However, to date no *MIC* genes have been identified in mouse or rat (Gunther and Walter 2001; Walter and Gunther 2000).

The MHC has also been identified in several non-mammalian species, including birds and bony and cartilaginous fishes. One of the most interesting differences in the genetic characteristics of the MHC occurs in teleost and avian species. In these species the class I antigen processing and transport genes: *TAP1* and *TAP2*, reside within the class I region and are found tightly linked with the class I loci (Clark et al. 2001; Flajnik and Kasahara 2001; Graser et al. 1998; Nonaka et al. 2001; Shiina et al. 2004b). This genetic organization is dramatically different from that found in mammals, where the class I antigen processing and transport genes are found within the class II region. These observations imply that the organization, which exists in avian and teleost species likely

represents the ancestral state and the location of the class I antigen processing and transport genes in the class II region of the mammalian MHC is likely a derived organization (Flajnik and Kasahara 2001).

Class II region

Comparative analysis of the class II genes in mammals identified five orthologous groups of class II genes (*DP*, *DM*, *DO*, *DQ*, *DR*) that are present only in mammalian species (Kumanovics et al. 2003). Phylogenetic analyses concluded that the alpha and beta genes of *DP*, *DO*, *DM*, *DQ*, and *DR* class II gene families originated around the time of the separation of placental mammals and marsupials (Takahashi et al. 2000). One notable exception is the presence of a *DM* gene in chicken; however, no other shared class II genes groups were identified in the chicken (Kaufman et al. 1999b). This conservation suggests that the *DM* gene group diverged early in the evolution of the MHC. While these five class II gene groups are conserved across mammals, there is considerable variation in the number of functional and non-functional (pseudogenes) alpha and beta subunit genes within each gene group due to species specific local duplications and deletions (Kumanovics 2002). For example, the two non-classical HLA class II gene pairs, *DM* and *DO*, each have a single functional gene pair encoding of the alpha and beta subunits (Beck and Trowsdale 1999; MHC Sequencing Consortium 1999), while the *DP*, *DQ*, and *DR* gene groups each have at least on functional set of alpha and beta genes and one to several pseudogenes. The *HLA-DP* gene group has two *DPA* and one *DPB* pseudogenes, *HLA-DQ* gene group has one *DQA* and two *DQB*

pseudogenes, and the *HLA-DR* gene group contains a variable number of *DRB* genes and pseudogenes (Beck and Trowsdale 1999; Kumanovics et al. 2003). The cat MHC exhibits the most dramatic variation in the class II genes. The class II region of the feline MHC shows remarkable conservation of gene organization with HLA in regard to the non-class II genes and non-classical class II genes but not the classical class II genes (Beck et al. 2001). In cats the *DQ* genes are absent and the *DP* gene group has been inactivated to pseudogenes. In contrast the *DR* genes have been expanded and reorganized resulting in three functional *DRA* genes, three functional *DRB* genes, and one *DRB* pseudogene (Beck et al. 2001; Yuhki et al. 2003).

A pair of divergent class II genes has been identified in the class IIb region of the cattle MHC (BoLA). In BoLA, a chromosomal inversion resulted in the disruption of the class II region, which translocated a portion of the class II region near the centromere (Andersson et al. 1988; Band et al. 1998; McShane et al. 2001; Skow et al. 1996). The class IIa region is contiguous with the class III and class I regions and contains the *DM*, *DQ*, and *DR* gene families, but no *DP* genes (Hess et al. 1999). The BoLA class IIb region contains two divergent class II genes named *DYA* and *DIB* (Stone and Muggli-Cockett 1990; van der Poel et al. 1990) in addition to the *LMP2*, *LMP7*, *TAP1*, *TAP2*, *DOB*, and *DOA* which are present in the class II regions of all mammalian MHCs studied to date (Hess et al. 1999). Comparative sequence analysis of *DYA* and *DIB* exons has not identified any orthologous genes in human or mouse. However, orthologs of the DYA gene have been identified in sheep (Wright and Ballingall 1994; Wright et al. 1994) and goats (Mann et al. 1993), but not in pigs indicating that the DYA gene

arose by duplication after the divergence of Suidae and Bovidae and is likely a member of a ruminant specific class II gene pair (Ballingall et al. 2001; Skow et al. 1996; Stone and Muggli-Cockett 1993). Additional research of the DY gene pair revealed a limited pattern of expression in peripheral dendritic cells and relatively little polymorphism in the region encoding the peptide binding region (Ballingall et al. 2001; Ballingall et al. 2004).

Orthology is unable to be determined in non-mammalian class II genes. In non-mammalian vertebrates the class II genes frequently appear to cluster in a species specific manner (Kumanovics et al. 2003; Shiina et al. 1999a; Shiina et al. 2004b). In avian species, phylogenetic analysis of the quail and chicken class II genes revealed that the quail genes are more closely related to each other than to chicken genes. These results indicate that the class II genes were duplicated after the speciation of quail and chicken and may explain the difference in the number of class II genes observed between chicken and the quail MHCs (Edwards et al. 1995; Shiina et al. 2004b). The quail MHC contains ten class II genes, seven class II ß genes and three non-classical class II genes, which is in stark contrast to the chicken's B locus that contains a single polymorphic class II ß gene (Kaufman et al. 1999b; Shiina et al. 2004b). Since the class II genes appear to have evolved independently within avian species the difference in the number of class II genes may simply reflect the amount of duplication and expansion that has occurred within each species (Shiina et al. 1999a; Shiina et al. 2004b).

Investigations of the class II region in teleost fishes have identified some characteristics that are unique to the teleost MHC. First, the class II region is not linked

to the class I region and in many instances they actually exist on separate chromosomes (Bingulac-Popovic et al. 1997; Michalova et al. 2000). In the zebrafish the class II loci are distributed among at least three different chromosomes and are non-syntenic with the class I loci (Kuroda et al. 2002). The number and location of the class I and class II genes also differs between bony fishes and mammals. In teleost, the class II region of the MHC is more flexible and susceptible to dramatic expansion than the class I region, which is opposite of what is observed in the mammalian MHC where the class I genes have expanded dramatically (Flajnik and Kasahara 2001).

Class III region

The class III region is the most gene dense region of the MHC. This small region contains a variety of genes involved in diverse fundamental biological processes. While the rest of the MHC is characterized by of gene families and pseudogenes, the class III region is relatively stable. In humans, the class III region contains 62 loci of which two are pseudogenes, this contrasts the HLA class I region contains 122 loci and 66 of them are pseudogenes (Shiina et al. 2004a).

Comparative analysis of the class III region in mammalian species, including human, mouse, rat, and pig has revealed a striking degree of conservation in gene order and gene content (Peelman et al. 1996; Walter et al. 2002). In contrast to class I and class II regions which generally share the same gene content, but are more prone to rearrangements (Peelman et al. 1996). Nonetheless, there is one area within the class III region where duplication events have occurred and it is centered on the *C4* gene

(Kumanovics et al. 2003); however, the exact duplication units are both species and haplotype specific (Yung Yu et al. 2000). Interestingly investigation of the rat class III region did not discover a tandem duplication of the *TNX-CYP21A2-C4-STK19* genes, however it did identify additional copies of *C4*, *CYP21A2*, and *STK19* in the class II and class III boundary (Walter et al. 2002). Investigations of other mammalian species such as the pig have not revealed any evidence of *C4* or *CYP21A2* duplication (Peelman et al. 1996).

In non-mammalian species the gene content and order of the class III region is not well conserved. In fact, *C4* is the only class III gene that has been identified in chicken and quail (Kaufman et al. 1999b; Shiina et al. 2004b) and to date *C4* and *BF* are the only class III genes identified in cartilaginous fish. Nonetheless, these class III genes are have been maintained in tight linkage with the class I and class II loci. Notably, conserved linkage is not observed in the teleost fish were the class III genes (*C4* and *BF*) are not linked to the class I and/or class II genes, which likely indicates that the organization in teleost fish is derived (Nonaka et al. 2001).

MHC evolution

The MHC has now been studied in a variety of vertebrate species from mammals to cartilaginous fish (Kulski et al. 2002). Comparative studies of the MHC have investigated the stability of gene content and gene order in the MHC over time (Kumanovics et al. 2003; Trowsdale 1995). MHC class I and class II genes are present in all jawed vertebrates, but absent in jawless vertebrates (Flajnik and Kasahara 2001;

Trowsdale 1995). Thus, the MHC originated in a common ancestor of jawed vertebrates (Kasahara et al. 2004). Moreover, the evolution of the MHC in jawed vertebrates is coincident with the emergence of the adaptive immune system, which has generated tremendous interest among the research community given the association between the MHC and disease (Kumanovics et al. 2003).

Evolution of the mammalian MHC has been studied in great detail via comparative analyses of both nucleotide sequence and gene mapping data generated for a variety of mammalian species. Conserved characteristics have been observed within the mammalian MHC thereby providing some insight into MHC evolution. There appears to be a remarkable degree of conservation of gene content and function within the MHC (Trowsdale 1995). Also, the class I and class II genes exist as members of gene families, and within these families there are both pseudogenes and functional genes (Klein et al. 1998; Yeager and Hughes 1999).

Comparative analyses of the MHC have revealed that all jawed vertebrates contain class I and class II genes, which appear to have similar conserved structure and function (Trowsdale 1995), but no shared class I gene lineages can be detected among mammals (Hughes and Nei 1989b). The class II gene groups demonstrably are orthologous only in mammals, except for the class II DM genes which have been found in the chicken MHC (Kaufman et al. 1999b). The multi-gene families found within the MHC arose out of gene duplication events that occurred at some point during the evolution of jawed vertebrates. Collectively, gene duplication is an important evolutionary process resulting in genetic variation, new gene function(s), unique

expression patterns, and divergence among species (Lynch and Conery 2000; Otto and Yong 2002). When gene duplication occurs the resulting daughter gene is considered the raw material from which novel genes may ultimately develop (Lynch and Conery 2000; Otto and Yong 2002). Following the duplication of a functional gene or genes, the daughter gene(s) may either diversify and acquire a novel function or be silenced by the accumulation of degenerative mutations (pseudogene) (Lynch and Conery 2000). The class I and class II regions of the human MHC both contain many pseudogenes, with approximately 50% of the class I HLA genes being non-functional pseudogenes (Shiina et al. 2004a). While many pseudogenes are not expressed and therefore have no biochemical function(s) they may play a role in the generation of new alleles through gene conversion (Grimsley et al. 1998).

These observations are best explained by two complementary models of evolution: the "birth and death model" (Nei et al. 1997) and the "accordion model" (Klein and Sato 1998; Klein et al. 1998). It was observed that the class I and class II genes within a single species are sometimes more closely related to the class I and class II genes for a different species (Nei et al. 1997). Thus, concerted evolution, which is believed to homogenize the gene sequences for member genes of a multi-gene family via gene conversion (interallelic recombination), could not explain the evolution of these gene families (Nei et al. 1997). Moreover, Nei (1997) proposed that the birth and death model of evolution best explains the evolution of multi-gene families within the MHC. Simply stated this model proposes that a single ancestral gene undergoes serial duplications or "births", producing multiple copies of a gene, which can subsequently be

modified by mutation. These newly formed genes can either be maintained in the population for long periods of time and become functionally divergent, or they may become non-functional pseudogenes which are referred to as "deaths" (Nei et al. 1997). Similarly, the accordion model proposes that because the number of class I and class II MHC genes varies among species, and the number of these genes does not steadily increase with increasing levels of divergence among species, then it is likely that the gene families undergo expansions by gene duplication followed by contractions caused by deletions or mutational inactivation of member genes (Klein and Sato 1998; Klein et al. 1998). Thus, the evolution of these gene families mimics the mechanism by which an accordion is played. Both models explain the evolution of gene families in the MHC via multiple rounds of gene duplication and subsequent modification. However, they do not explain the high levels of polymorphism that are maintained in the functional class I and class II genes.

MHC polymorphism

The classical class I and class II genes of the MHC display an extremely high degree of polymorphism within the nucleotide sequence that encodes the peptide binding region (PBR) of the respective proteins (Hughes and Hughes 1995; Hughes et al. 1994; Hughes and Nei 1988; Hughes and Nei 1989b). Initial explanations of this high degree of polymorphism were based on the biological function of the MHC molecules. It was proposed that the high degree of polymorphism seen in the PBR of class I and class II genes was the result of overdominant selection. Importantly, the maintenance of

overdominant selection favoring heterozygosity within the PBR would potentially allow the organism to bind and present a greater range of antigenic peptides to T cells (Doherty and Zinkernagel 1975). Thus, heterozygosity within the PBR could potentially protect the organism from a wider spectrum of pathogenic organisms (Doherty and Zinkernagel 1975). Additional investigations of the nucleotide sequences of class I and class II genes revealed an increased rate of non-synonymous to synonymous nucleotide substitution for codons encoding the PBR of the class I and class II molecules (Hughes and Nei 1988; Hughes and Nei 1989b; Hughes and Yeager 1998a; Hughes and Yeager 1998b; Yeager and Hughes 1999). Moreover, because synonymous substitution is largely selectively neutral, instances where the non-synonymous substitution rate is considerably greater than the synonymous substitution rate are often indicative of balancing selection. Thus, high levels of non-synonymous polymorphism routinely observed within the classical class I and class II genes is hypothesized to be the result of a form of balancing selection (overdominant selection) acting to maintain heterozygosity within the PBR (Hughes and Nei 1989b; Hughes and Yeager 1998a; Yeager and Hughes 1999).

Study objectives

To date the MHC has been characterized with respect to its genomic organization and gene content in a variety of different vertebrate species. Some species have complete genomic sequence available for the MHC, while the MHC of other species has been analyzed by partial sequencing, radiation hybrid mapping, or linkage analysis. To date there has been limited characterization of the horse MHC (ELA). The primary focus of

these investigations has been the identification of ELA class I and class II genes and the subsequent sequence analysis for polymorphisms. Thus, this study seeks to undertake a comprehensive characterization of the entire ELA complex.

Our first objective in this study is to generate a sequence-ready bacterial artificial chromosome (BAC) contig that spans the horse MHC, which is referred to as the equine lymphocyte antigen (ELA) complex. This BAC contig will serve as a template for the characterization of ELA which can be performed in a variety of manners. The second objective of this study is to generate a physical map of ELA that contains a portion of the gene content for the ELA genomic region. The identification of genes within the MHC of the horse provides the opportunity to perform comparative analyses, which is the focus of the final objective in this study. The final objective of this study is to perform comparative analyses of the gene content and organization of ELA with respect to the MHC region of other mammalian species. We propose that the characterization of the genomic organization and gene content of ELA will provide insight into the evolutionary relationships related to disease, polymorphism, and linkage disequilibrium in this region.

CHAPTER II

CONSTRUCTION OF BACTERIAL ARTIFICIAL CHROMOSOME (BAC) CONTIGS FOR THE ELA COMPLEX*

Introduction

The major histocompatibility complex (MHC) is the most gene dense region of the human genome (MHC Sequencing Consortium 1999) and contains many genes that function in the innate and adaptive immune responses (Trowsdale 2001). By convention, the MHC is divided into three regions (class I, class II, and class III) based upon gene function. The definitive genes of the MHC are the class I and class II genes, which encode cell surface glycoproteins that present endogenous and exogenous peptides to lymphocytes, respectively. Peptide-loaded MHC class I molecules are recognized by the antigen-specific receptors of CD8+ cytotoxic T lymphocytes and are expressed on virtually all nucleated mammalian cell types, with the notable exception of most forms of trophoblasts (Donaldson et al. 1992). Peptide-loaded MHC class II molecules have a more limited expression profile that is closely associated with, but not restricted to, antigen presenting cells. In the horse, MHC class II molecules are expressed constitutively on T lymphocytes (Crepaldi et al. 1986), an expression pattern that differs from that of humans and mice. MHC class II molecules are recognized by the

antigen receptors of CD4+ helper T lymphocytes. The class III region is highly gene dense, containing a number of genes that function in the immune response such as complement component 4 (*C4*) and tumor necrosis factor alpha (*TNF*), and other genes apparently not functionally associated with the immune system.

MHC-encoded genes have been associated with susceptibility to numerous infectious and non-infectious diseases in humans, such as arthritis, diabetes, psoriasis, making the MHC a region of considerable biomedical interest (Choonhakarn et al. 2002; Dawkins et al. 1999; Dawkins et al. 1983; Nishimura et al. 2003; Okamoto et al. 2003). However, association of diseases with the MHC is complex and identification of contributing genes is difficult due to high levels of polymorphism, linkage disequilibrium, and gene duplications (Apanius et al. 1997). Insight into the function and evolution of the MHC can be gained from comparative mapping in various mammalian species.

The genetic content of the MHC was first described in mice and humans, but has since been characterized in other mammalian species and appears to be evolutionarily conserved (Beck et al. 2001; Chardon et al. 1999a; Gunther and Walter 2001; McShane et al. 2001; Walter et al. 2002). However, significant differences in the number and physical organization of the genes within the MHC have been demonstrated in several species of the Artiodactyla (Band et al. 1998; Chardon et al. 1999a; Smith et al. 1995). Gene mapping and complete sequence information has been published for the MHC of several species such as human (MHC Sequencing Consortium 1999), mouse (Amadou et al. 1999), rat (Gunther and Walter 2001; Walter et al. 2002), pig (Chardon et al. 1999a),

cat (Beck et al. 2001; Kuwahara et al. 2000), and cattle (Ellis et al. 1999; Lewin et al. 1999), which has provided data that has been useful in delineating the evolutionary development of the MHC. Given the demonstrated value of comparative studies as well as the limited information that currently exists for the equine MHC (ELA), we feel that significant information relevant to the evolution of the MHC can be derived from a comprehensive analysis of ELA.

Notable differences in the physical organization of the genes within the MHC have been demonstrated in some species. The swine leukocyte antigen (SLA) is disrupted by the centromere, which separates the class II region from the rest of the MHC (Smith et al. 1995). Another example of a deviation in the gross genomic organization of the MHC occurs in the bovine leukocyte antigen (BoLA) complex. A large portion of the class II region was translocated near the centromere by a large chromosomal inversion (Band et al. 1998; Muggli-Cockett and Stone 1988; Skow et al. 1996). Additional FISH analysis demonstrated that the inversion of gene order is present in all ruminants examined to date, and that the organization of the cattle MHC is likely an ancestral condition for all ruminants (McShane et al. 2001).

The MHC of the horse (*Equus caballus*), which is referred to as ELA for equine lymphocyte antigen (Bailey et al. 2000; Marti et al. 1996), is located on ECA20q14→q22 (Ansari et al. 1988; Makinen et al. 1989). A number of diseases such as equine infectious anemia virus (McGuire et al. 2000; McGuire et al. 1997), equine squamous cell carcinoma (Perez et al. 1999), uveitis (Romeike et al. 1998), sarcoids

(Brostrom et al. 1988; Chambers et al. 2003; Goodrich et al. 1998; Meredith et al. 1986), mold allergy (Curik et al. 2003), and sweet itch (Kurotaki et al. 2000) have been associated with ELA. To date, much of the research on ELA has been targeted on the identification and characterization the class I and class II genes. Restriction fragment length polymorphism (RFLP) analysis indicated that 20–30 class I genes and/or pseudo genes might be present in the ELA (Alexander et al. 1987), but only one complete genomic sequence and four or five full-length cDNA sequences of ELA class I genes have been reported to date (Barbis et al. 1994; Carpenter et al. 2001; Ellis et al. 1995). Analysis of the class II *DQA* gene revealed one highly polymorphic locus (Fraser and Bailey 1998), while single strand conformation polymorphism (SSCP) analysis of the classical class II genes identified two transcribed *DQB* loci in the domestic horse (Horin and Matiasovic 2002). Investigation of class II *DRA* gene yielded a single polymorphic locus (Albright-Fraser et al. 1996). While investigation of the class II *DRB* gene indicated that three *DRB* loci existed in the domestic horse (Fraser and Bailey 1996), while recent studies found some Andalusian stallions (an ancient breed of domestic horse) and every Przewalski's horse investigated appear to have at most two *DRB* genes (Hedrick et al. 1999). Currently, it is not known if the number of *DRB* loci vary between MHC haplotypes, as is found in other species (Ellis and Ballingall 1999; Gongora et al. 1997).

In order to better define the genomic structure of ELA, we used a comparative genomics approach to identify BAC clones containing equine MHC genes and then

performed a series of analyses to assemble the BAC clones into contigs of defined gene content and chromosomal orientation. We propose that characterization of ELA will provide insight into the evolutionary relationships related to disease, polymorphism, linkage disequilibrium, and recombination in this region.

Materials and methods

Design of overlapping oligonucleotide (overgo) probes

Using the available human genomic sequence for HLA as a framework, genes were chosen approximately every 150 kb for overgo primer design. Highly conserved regions within the targeted genes were identified by aligning available gene sequences from multiple species using GenBank's Pairwise BLAST (Tatusova and Madden 1999). The conserved sequence regions were analyzed using RepeatMasker (http://www.repeatmasker.org) to identify the presence of any repetitive elements. Overgo primers were designed to the conserved, non-repetitive regions within the genes using the Overgo Maker program (http://www.genome.wustl.edu/tools/?overgo.html). The designed overgo primers were screened against GenBank using BLAST (Altschul et al. 1990) to confirm gene specificity and exclude the presence of repetitive elements.

Overgo probe labeling and filter hybridization

The overgo primers were radioactively labeled using a modification of a previously described protocol (http://genome.wustl.edu/tools/protocols/mapping/Prehyb.pdf). A 10

μL reaction containing 1 μM forward overgo primer and 1 μM reverse overgo primer, 150 Ci/mmol each of ^{32}P dATP and ^{32}P dCTP (Amersham Biosciences, Piscataway, NJ), 2 U Klenow fragment DNA polymerase (Roche, Indianapolis, IN), and 1x DNA Polymerase Buffer (Promega, Madison, WI) was incubated at 37°C for 30 minutes. For fill-in labeling, 1 μL of a 250 μM dATP and dCTP mixture was added to each reaction and incubated at 37°C for 30 minutes (Han et al. 2000). Unincorporated nucleotides were eliminated using Sephadex G-10 gravity flow columns.

The equine *(Equus caballus)* CHORI-241 BAC library was constructed by the Children's Hospital Oakland Research Institute (http://bacpac.chori.org/equine241.htm). The library was constructed by size fractionating partially *EcoR*I digested DNA from peripheral blood neutrophil cells of a thoroughbred male and cloning the fragments in the pTARBAC2.1 vector. The CHORI-241 library has an 11.8 fold genomic representation with approximately 190,652 recombinant clones which have an average insert size of 171 kb (http://bacpac.chori.org/equine241.htm). The BAC clones were arrayed into 384 well plates and high density filters were prepared from the 384 well plates.

High density filters from the CHORI-241 equine BAC library (http://bacpac.chori.org/equine241.htm) were screened by pooled overgo hybridization. The labeled overgo probes were pooled in equal concentrations and added to the hybridization solution (20x SSPE, 10% SDS, 5% milk, and 100 x Denhardt's) containing 50% formamide. The hybridization solution containing the pooled overgo probes was

incubated in a boiling water bath to denature the overgo probes, quenched on ice, and added to the filters of the BAC library. Hybridization was performed at 42°C for 15-17 hours. Following overnight hybridization the filters were washed three times at 55°C for 15 minutes in 2x SSPE. The filters were exposed to film over intensifying screens for two to seven days at -80°C and the autoradiograms were subsequently developed. The autoradiograms were analyzed by hand to identify to the positive BAC clones.

Identification of gene content by hybridization

Secondary filters of the positive BAC clones were created to allow for individual hybridization of radioactively labeled gene specific overgo probes to these filters. The radioactive labeling and hybridization protocols were the same as previously described. These hybridizations were utilized to identify which BAC clones were positive for each overgo probe. Gene content was determined by a positive hybridization of a gene specific overgo probe to a BAC clone.

BAC fingerprinting, Southern blot analysis and contig assembly

DNA was extracted from all of the MHC positive BAC clones using the Qiagen Midi Prep Kit (Qiagen, Valencia, CA). DNA was quantified on 0.75% TBE agarose gel containing ethidium bromide. BAC fingerprinting and contig assembly was performed using a modification of the Marra et al. protocol (1997). For each BAC clone 1 μg of DNA was restriction digested with *Bam*HI. The restriction fragment patterns of each

BAC clone were analyzed by running each BAC clone's restriction digestion products on a 0.65% TBE agarose gel containing ethidium bromide, which was run at 50 volts for 20 to 24 hours at 4°C. The gel images were digitally captured with the Alpha Innotech ChemiImager system and sizes were assigned to each restriction fragment pattern using IMAGE 3.0 software (Sulston et al. 1988). Preliminary contig assembly was performed using FPC V4.7.9 (Soderlund et al. 1997).

Each DNA fingerprinting gel was Southern blotted onto nylon filters for subsequent hybridizations. Southern blot filters from the BAC fingerprinting experiments were hybridized with individual gene-specific overgo probes to confirm the gene content and to identify the number of class II genes (*DQA*, *DQB*, *DRA*, and *DRB*) present in the BAC clones. The overgo probe labeling and hybridization protocols were described previously.

BAC end sequencing

Sequencing primers were designed to flank the insertion site of the CHORI-241 BAC vector, pTARBAC2.1. The T7.29 primer (5'-GCCGCTAATACGACTCACTATAGGGAGAG) was obtained from the CHORI website (http://bacpac.chori.org/cyclesere.htm) and the SP6.26 primer (5'-CCGTCGACATTTAGGTGACACTATAG) was designed from the published sequence of pTARBAC2.1. Each BAC clone was end sequenced on an ABI 3100 automated capillary sequencer using the following protocol. The 10 μL sequencing reaction

contains 1 μg BAC DNA, 2 μL Big Dye (Applied Biosystems, Foster City, CA), 2 μL half Big Dye (Genetix, Boston, MA), 0.5 μL of Master Amp PCR Enhancer (Epicentre, Madison, WI), 1 μL of T7.29 or SP6.26 sequencing primer (10 μM), and bring to 10 μL final volume with autoclaved double distilled water. The thermal profile is as follows: 2 min at 96°C; 60 cycles of 30 s at 96°C, 30 s at 50°C, 4 min at 60°C; 15 min at 65°C. The unincorporated nucleotides were removed from the sequencing reaction with Spin-50 spin columns (BioMax, Odenton, MD) and the reaction was dried down in preparation to be sequenced on the ABI 3100.

End sequence specific polymerase chain reaction (PCR) and overgo hybridization

A subset of ELA BAC clones that had been assembled into preliminary contigs by DNA fingerprinting or were predicted based upon their gene content to exist in the gaps between the contigs were chosen to have PCR primers designed from their unique end sequences. The end sequence specific PCR primers can also be referred to as sequence tagged site (STS) markers. The end sequences were analyzed using GenBank's BLASTn (Altschul et al. 1990) to identify homologies with published sequences both repetitive and gene specific. Repetitive sequences were masked to be avoided during primer design and PCR primers were designed to unique regions of the end sequences using Primer3 software (Rozen and Skaletsky 2000) (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Designed PCR primer pairs were screened using

RepeatMasker (http://www.repeatmasker.org) and GenBank's BLASTn (Altschul et al. 1990) to confirm specificity.

The end sequence specific PCR conditions for each primer pair were optimized using template DNA from the BAC clone the end sequence originated from. PCR was carried out in 25 μl reactions containing 50 ng of BAC DNA from individual clones as the template, 0.25 U of JumpStart Red Taq Polymerase (Sigma, St.Louis, MO), 0.8 mM dNTPs, 0.4 μM of each primer, Master Amp PCR Enhancer (Epicentre, Madison, WI), and 10× reaction buffer (100 mM Tris-HCl, 500 mM KCl, 15 mM $MgCl_2$, and 0.01% gelatin). The thermal profile was as follows: 2 min at 95°C; four cycles of 30 s at 95°C, 30 s at 58°C (–1°C/cycle), 25 s at 55°C; 30 cycles of 30 s at 95°C, 30 s at 54°C, 25 s at 55°C; 10 min at 65°C for most primer pairs, however there was some variation in the annealing temperatures.

End sequence specific overgo probes were designed in collaboration with Cornell University to confirm overlaps in regions where the repetitive nature of the sequence did not allow PCR primer pairs to be designed. Overgo design, radioactive labeling, and hybridization were performed using the previously described protocols. End sequence specific overgo primers were designed, radioactively labeled, and individually hybridized to dot blot filters containing BAC clones that were preliminarily assembled into a minimum tiling path of the contigs as previously described.

Fluorescence in situ hybridization (FISH)

Selected BAC clones (135 M23, 147 K21, 163 J11, 288 J19, 359 L18, 372 F10, 382 H22, 407 K07, 431 P04, 440 J07, 441 N13, 455 C07, 464 F03, and 528 E24) were labeled with biotin and/or digoxigenin according to manufacturer's instructions using BioNick Labeling System (Invitrogen, Carlsbad, CA) and DIG-Nick Translation Mix (Roche, Indianapolis, IN), respectively. The labeled BAC DNA was hybridized to horse metaphase chromosomes individually to confirm their location to ECA20q14→q22. BAC clones were also differentially labeled and co-hybridized to metaphase chromosomes to determine the relative positions of clones. DNA labeling, in situ hybridization, signal detection, microscopy, and image analysis were carried out as described in Raudsepp et al. (1999) and Chowdhary et al. (2003).

Results

Identification of BAC clones containing ELA genes

Twenty three overgo primer pairs were designed to conserved regions within exons of 22 orthologous genes found in the MHC (Table 1). Initial pooled screening of the equine CHORI-241 BAC library for 22 MHC genes yielded 103 positive BAC clones (Appendix A). The gene content of the positive BAC clones was determined by individual hybridization of gene-specific overgo probes to dot blot filters of the positive BAC clones. Every gene-specific overgo probe identified at least one BAC clone and in some cases multiple genes were found on a single BAC clone.

Fingerprinting and Southern blot analysis

Each BAC clone was fingerprinted by *Bam*HI digestion and digests were analyzed by agarose gel electrophoresis (Fig. 1 and Fig. 2) BAC clones were grouped based upon their gene content as predicted by overgo hybridization to dot blot filters to facilitate fingerprint analysis. Analysis of the fingerprinting gels by IMAGE 3.0 (Sulston et al. 1988) and FPC V4.7.9 (Soderlund et al. 1997) software resulted in the assembly of seven provisional contigs. The provisional contigs were assembled solely on the basis of the BAC fingerprint analysis without regard to predicted gene content.

Restriction fragment patterns generated from the BAC restriction digestion fingerprinting experiments were sized using the IMAGE3.0 (Sulston et al. 1988). Based upon the fragment sized assignments the approximate BAC insert sizes were estimated for each clone.

Table 1 Overgo probes utilized in the initial screening of the equine CHORI-241 BAC library.

Gene	Gene Product	Accession No. (species)	Forward Overgo Primer	Reverse Overgo Primer
BAT3	HLA-B associated transcript-3	13111924 (hs), 6114859 (mm), unpublished seq (bt)	CTGTTCATGACCGGAATGCCAACA	CCAACCATGACATAGCTGTTGGCA
BRD2	Bromodomain containing protein 2	12408641 (hs), unpublished seq (bt)	CCCATGAGTTACGATGAGAAGCGG	TGTCCAGGCTCAACTGCCGCTTCT
COL11A2	Alpha 2 subunit of type XI collagen	1480743 (ec)	CGCAAGAACCCTGCTCGCACCTGC	AGAGTTTCAGGTCCCGGCAGGTGC
DMB	MHC class II, DM beta	6636790 (ec)	TTAGCAACTTGGGGGAGCTCATTC	TACAGAATGTTCCCAGGAATGAGC
DOB	MHC class II, DO beta	32271 (hs), 561974 (mm), unpublished seq (bt)	GCAAAGGCTGACTGTTACTTCACC	CCTTTTCTGTCCCATTGGTGAAGT
DQA	MHC class II, DQ alpha	2226318 (ec)	ACTGAGAGAAGTGGCTACGGCAAA	GACTTCCAAGTTGTGTTTTGCCGT
DQB	MHC class II, DQ beta	164203 (ec)	ACATTGCCGAGTACTGGAACGGAC	TCCAGGACGTCCTTCTGTCCGTTC
DRA	MHC class II, DR alpha	976200 (ec)	GAGATTTTTCACGTGGATATGGAC	AGACCGTCTCCTTCTTGTCCATAT
DRB	MHC class II, DR beta	1228844 (ec)	ACGCCGAGTACTGGAACGGGCAGA	TCATCCAGGACGTCCTTCTGCCCG
ELA class I	MHC class I alpha chain	435020 (ec)	1) GCTTCATCACCGTCGGCTACGTGG 2) CCCAACACTGACCTTTGTGCTTTC	1) ACGAACTGCGTGTCGTCCACGTAG 2) AGGACATTAGATCAGGGAAAGCAC
FLOT1	Flotillin 1	6031143 (hs), 2149603 (mm)	TGTTTTTCACTTGTGGCCCAAATG	GAGACCACCATGGCCTCATTTGGG
GNL1	Guanine binding protein-like 1	807999 (hs), 311933 (mm), unpublished seq (bt)	AGCCTGGGCAGAGAAACGTGGTTA	GGCCTTGGCTGTCTTGTAACCACG

Table 1 Continued

Gene	Gene Product	Accession No. (species)	Forward Overgo Primer	Reverse Overgo Primer
HSPA1A	Heat shock 70 kDa protein 1A	5579469 (hs), 497937 (bt)	TGTCCATCCTGACGATCGACGACG	TTCACCTCGAAGATGCCGTCGTCG
MOG	Myelin oligodendrocyte glycoprotein	13645153 (hs), 548189 (mm), unpublished seq (bt)	CAGGATCCGGAATGTGAGGTTCTC	AAAACCTCCTTCATCTGAGAACCT
PBX2	Pre-B-cell leukemia transcription factor 2	35312 (hs), 2432012 (mm), unpublished seq (bt)	CATCGAACACTCGGACATCGCAG	GATCTGGGCAAGTTTGCTGCGATA
POU5F1	POU domain, class 5, transcription factor 1	17464418 (hs), 200117 (mm), 4103379 (bt)	CGTGAGGATTTTGAGGCTGCTGGG	CCCCTGCGAAAGGAGACCCAGCAG
PSMB8	Proteasome beta 8 subunit	38480 (hs), 405775 (mm), 3242945 (bt)	GTGATTGAGATTAACCCTTACCTG	CAGACATGGTGCCAAGCAGGTAAG
RXRB	Retinoid X receptor beta	30447 (hs), 3811374 (mm), unpublished seq (bt)	CTGGCCCCCCTGAAGATGTGAAGC	ACCCCTAAGACTGGTGGCTTCACA
TCF19	Transcription factor 19	833832 (hs), unpublished seq (bt)	ACTTTGGTCAATAATGTCCGACTC	GCCTGTGACCCCTTGGGAGTCGGA
TNF	Tumor necrosis factor alpha	6577092 (ec)	TACCTCATCTACTCCCAGGTCCTC	AGCCTTGGCCTTTGAAGAGGACCT
TRIM10	Tripartite motif-containing 10	12407412 (hs), 4731627 (mm), unpublished seq (bt)	AGTGTCTTGAGTGTCTAAGAAAAG	TGAATCTCCTCTCTCTCTTTTCTT
TRIM26	Tripartite motif-containing 26	16445440 (hs), 12275879 (mm), unpublished seq (bt)	AATCCTGAACCACCTGAGTACCCT	GTCTCTGTCTCTCCTTAGGGTACT

Species names are abbreviated as followed: human (hs), mouse (mm), cattle (bt), and horse (ec).

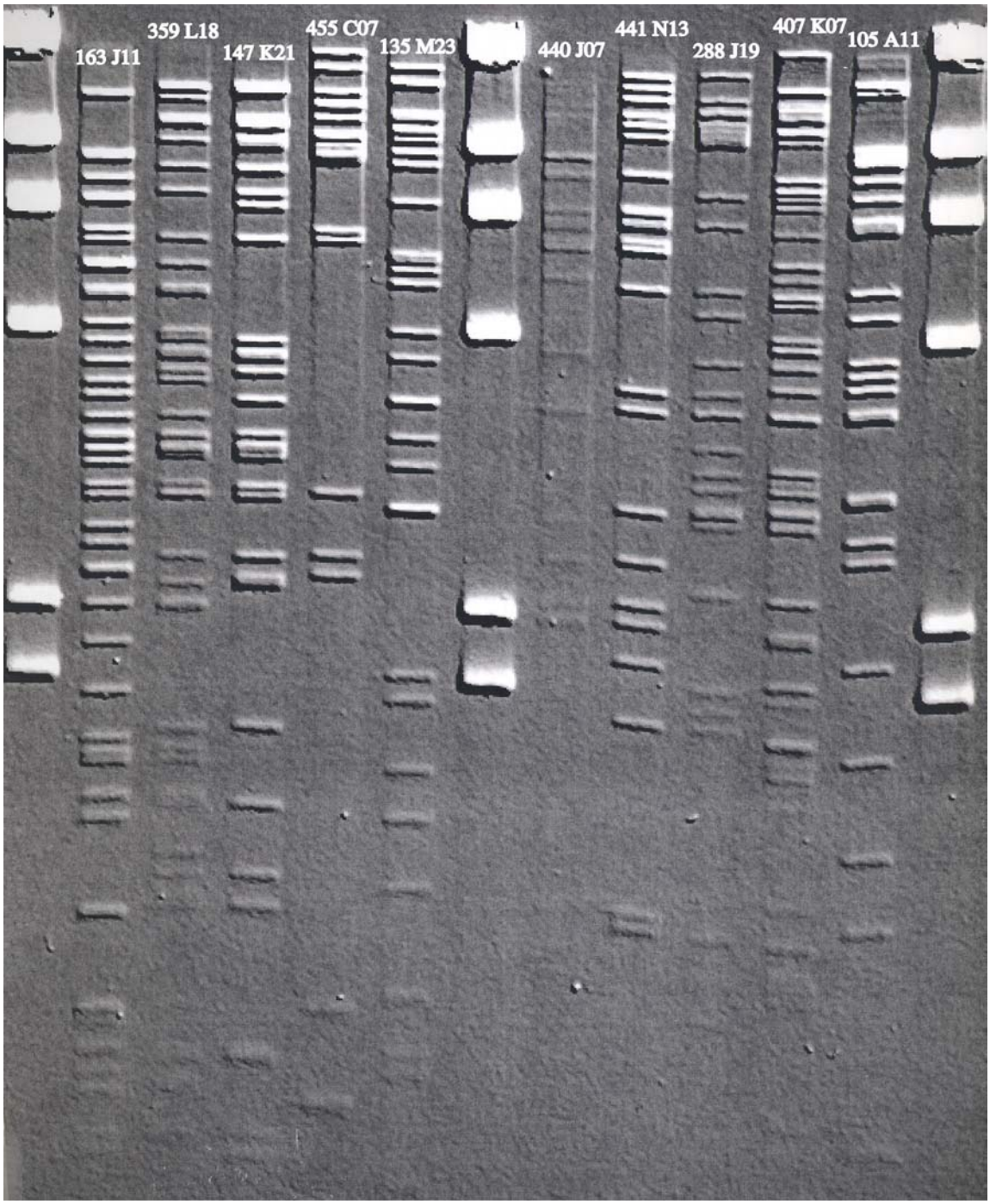

Fig. 1 Class II minimum tiling path BAC fingerprinting gel. The name of each BAC clone digested is listed above each lane. The markers on this gel are lambda DNA *Hin*d III digest.

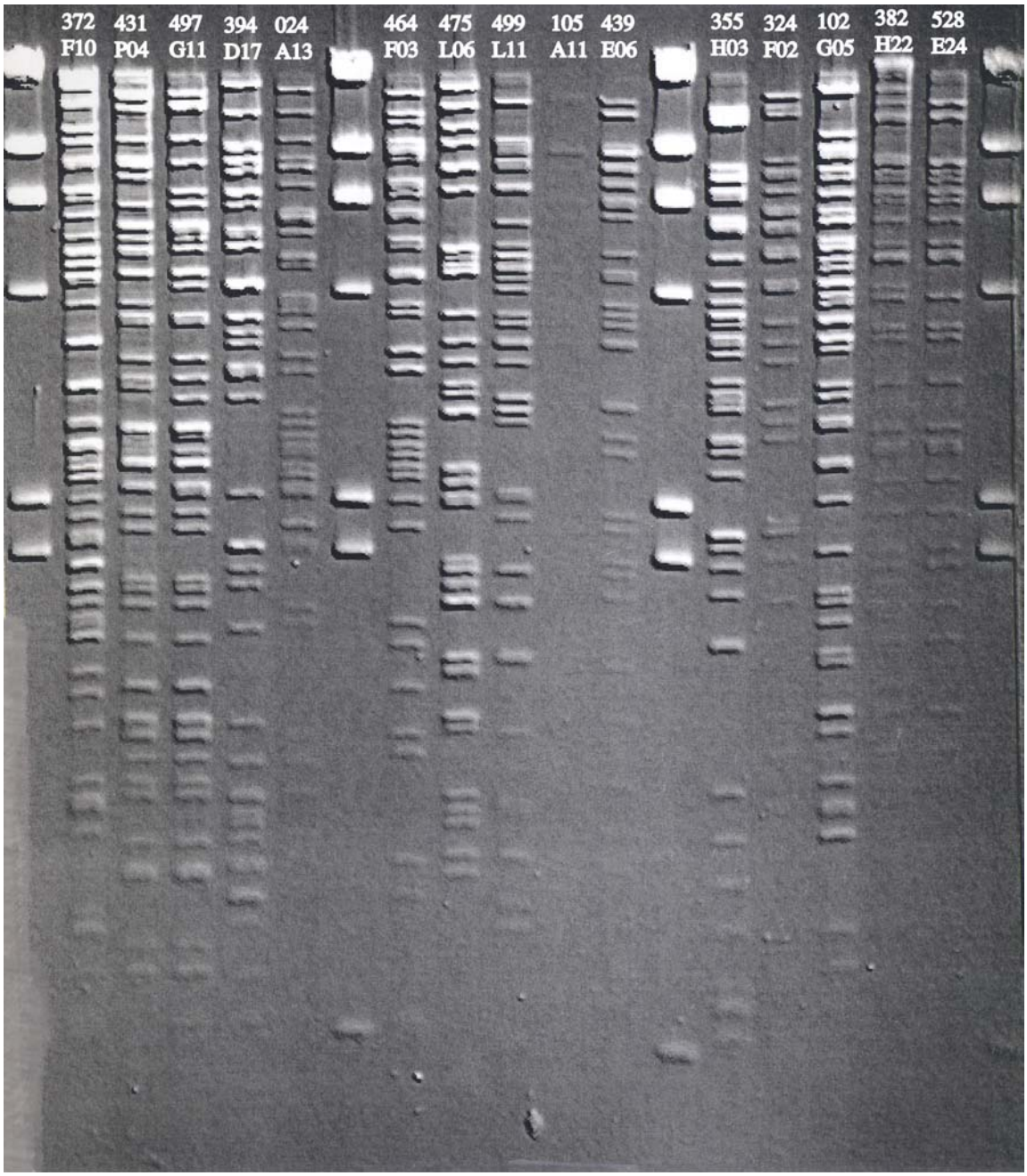

Fig. 2 Class III and I minimum tiling path BAC fingerprinting gel. The name of each BAC clone digested is listed above each lane. The markers on this gel are lambda DNA *Hin*d III digest.

Each fingerprinting gel was Southern blotted onto nylon filters for further analysis of gene content, which facilitated assembly of the BAC clones into contigs. The Southern blots were individually hybridized with gene-specific overgo probes to confirm the gene content predicted by hybridization of the dot blot filters and to identify the number of copies of class I and class II (*DQA*, *DQB*, *DRA*, and *DRB*) genes. Southern blot analysis of the class II BAC clones confirmed the existence of at least one locus for *ELA-DRA* and *ELA-DQA*, at least two *ELA-DQB* loci, and at least three *ELA-DRB* loci (Fig. 3). Southern blot hybridization confirmed the gene content of all BAC clones. Investigation into the potential duplicated segment in the class III region, which contains the *C4* and *CYP21A2* gene, was performed by Southern blot analysis with *C4* and *CYP21A2* overgo probes. The Southern blot analysis of the fingerprinted class III BAC clones identified only one positive fragment (Fig. 4A and Fig. 4B), which indicates that there is not a duplication of the *C4-CYP21A2* region in ELA.

Based upon the Southern blot confirmed gene content and fingerprinting results seven preliminary contigs were assembled for ELA. There were three contigs for the class II region: contig 1 contained *RXRB*, *BRD2*, and *PSMB8*; contig 2 contained *DOB DQB*, and *DRB*; and contig 3 contained *DQA*, *DQB*, *DRA*, and *DRB*. There was one contig for the class III region. There was three contigs for the class I region: contig 1 contained *POU5F1*, *TCF19*, *FLOT1*, and *GNL1*; contig 2 contained *TRIM10*, *TRIM26*, and class I genes; and contig 3 contained *MOG*.

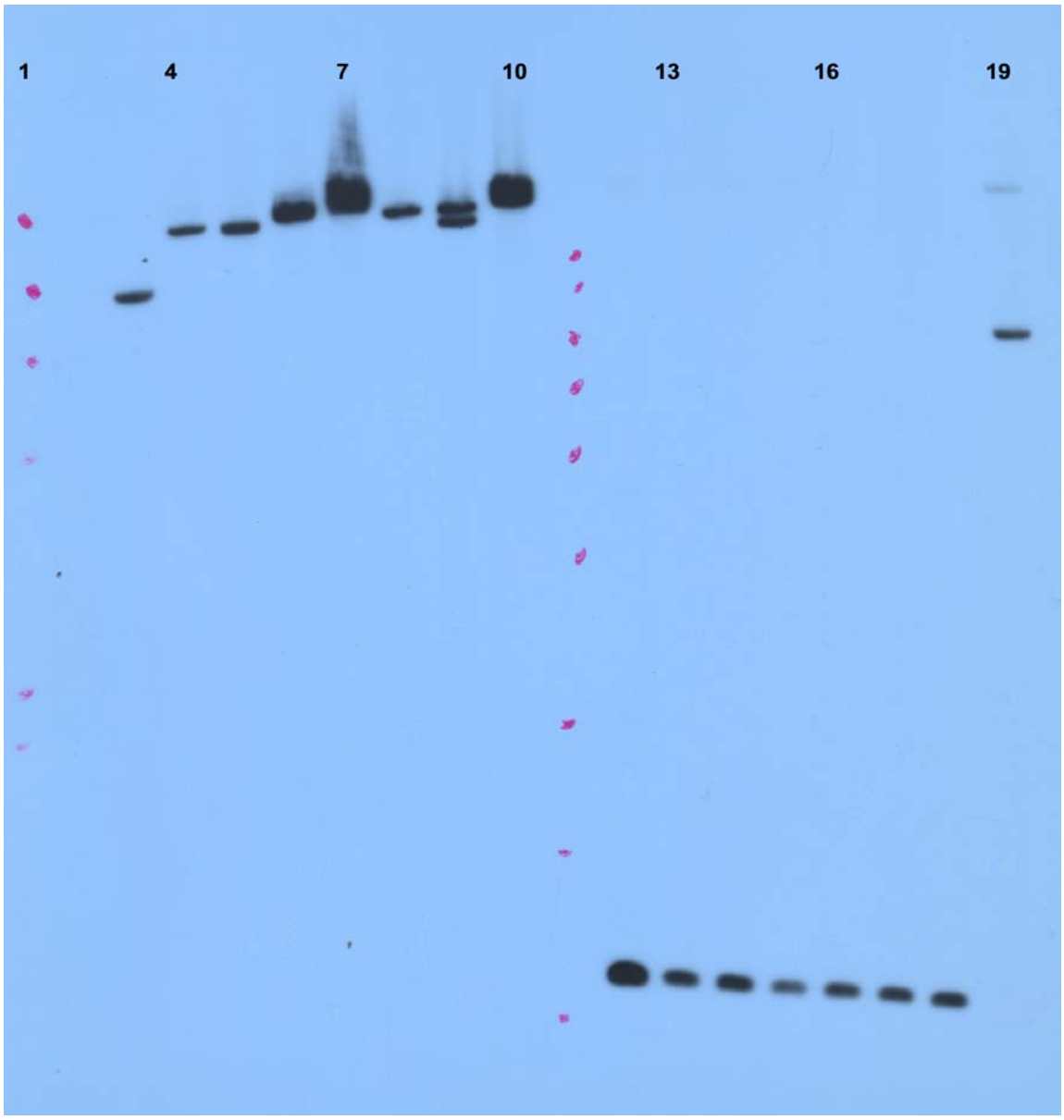

Fig. 3 Southern blot hybridization with the *DQB* overgo probe of the BAC fingerprinting filter. Starting at lane 1 the samples are as follows: lambda *Hind* III marker, 492 I22, 528 G20, 008 K14, 132 I24, 136 H02, 163 M01, 168 B14, 189 L11, 400 G19, 1 kilobase ladder, 012 F15, 013 C16, 158 E20, 363 O17, 345 J21, 373 P04, 498 L17, 389 G09.

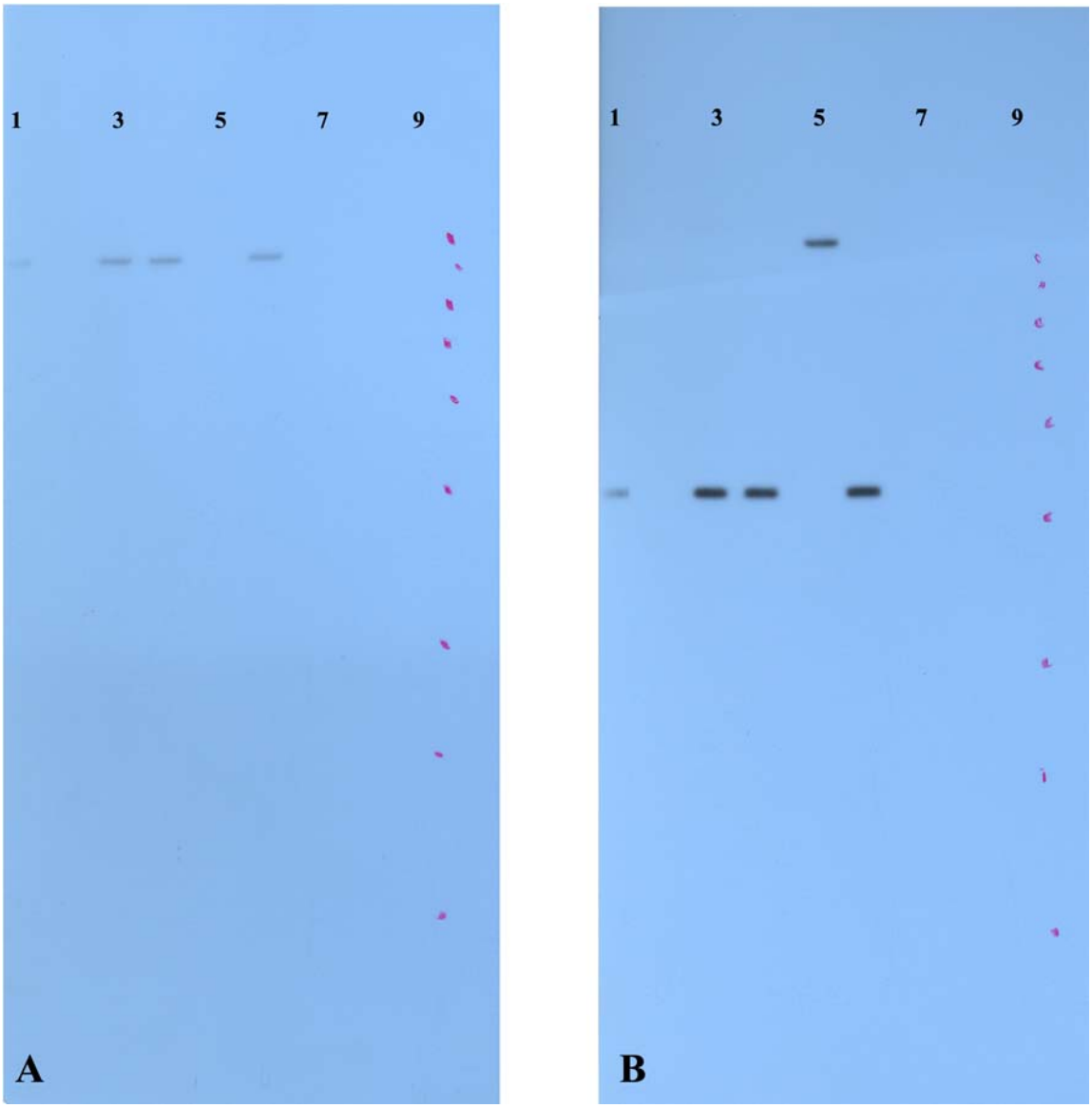

Fig. 4 Southern blot hybridization of BAC fingerprinting gel filters with gene specific overgo probes. The filter for ELA BAC digest 4 was used in these hybridizations. The samples starting with lane 1 are as follows: 372 F10, 522 P24, 348 K19, 341 J05, 164 I22, 006 E13, 406 C23, 510 G12, 169 B13, and one kilobase ladder. **A** Southern blot hybridization was performed with a *C4* overgo probe. **B** Southern blot hybridization was performed with a *CYP21A2* overgo.

End sequencing and analysis

All 103 BAC clones were end sequenced from both ends using modified T7 and SP6 primers. End sequence was obtained from both the T7 and SP6 end of 75 BAC clones. In some cases end sequences could not be obtained from one end of the insert despite multiple attempts, usually SP6 was the problematic side of the insert to sequence. Twenty one of the BAC clones were only sequenced from the T7 end and one BAC clone was sequenced from just the SP6 end. Six BAC clones (004 H14, 022 H20, 069 N18, 063 F17, 337 M04, 371 F04) were unable to be end sequenced and were subsequently dropped from the analysis.

The end sequences that were obtained were analyzed to identify homologies with published sequences for both genes and repetitive elements. The majority of the end sequences contained non-coding sequences and/or repetitive elements. However, 10 genes within the MHC were identified in the end sequences: *APOM*, *C2*, *CYP21A2*, *LY6G6C*, *MOG*, *MSH5*, *NRM*, *PPP1R10*, *RDBP*, and *TAP2*. All of the gene sequences, except *MOG* and *TAP2* were used to design additional gene-specific overgo probes (Table 2).

Table 2 End sequence derived gene specific overgo primers.

Gene	Gene Product	Accession No.	Forward Overgo Primer	Reverse Overgo Primer
APOM	apolipoprotein M	CC875245	CTAGTATCCCCACTCACTGTAGAG	TTACCAGCGCTTCCTCCTCTACAG
C2	complement component 2	CC875225	AATCCATGACTCCTGCATGGCATG	CACCCCAGATTGTATGCATGCCAT
CYP21A2	cytochrome P450, family 21, subfamily A, polypeptide 2	CC875248	1) GCTTCTAGGGAATTCTCTTTCCTC 2) ATTCTCTCTCCTCACCTGCAGCAT	1) AGATGATGCTGCAGGTGAGGAAAG 2) GGTGAGGTAACAGATGATGCTGCA
LY6G6C	lymphocyte antigen 6 complex, locus G6C	CC875223	CACTTGCAGAAATGGCTTGGCACT	TTTCTTGGCAGCCTCAAGTGCCAA
MSH5	mutS homolog 5 (E. coli)	CC875243	ATTCATGGTTCTGGCCCCACCTCT	TCTCAAGCTTCTCCAGAGAGGTGG
NRM	nurim (nuclear envelope membrane protein)	CC875241	AGGCCCAAATCCATGAACTGAAGG	AATCAGAGGGACAGCCCCTTCAGT
PPP1R10	protein phosphatase 1, regulatory subunit 10	CC875249	TAGGGATTCTCCTGCACATGCTAC	AAACTCAAGGGACCAGGTAGCATG
RDBP	RD RNA binding protein	CC875221	AGCCTCTGGTATACCATGGTTTGG	GGGAAACCCAATCTATCCAAACCA

Two overgo primer set was designed for the CYP21A2 gene and they are denoted by numbers 1 and 2.

End sequence specific PCR and contig assembly

BAC clones were chosen for further analysis based upon their gene content, position within the preliminary contigs, and quality of end sequence. The selected BAC clones non-coding, non-repetitive end sequences were used to design end sequence specific PCR primers, these are also called sequence tagged site (STS) PCR primers (Table 3). Overlapping BAC clones were confirmed by PCR amplification using BAC clones as templates. If an end sequence specific PCR primer set generated an amplicon in a particular BAC clone, then the BAC from which the primer set was designed is determined to overlap with the template BAC clone (Fig. 5). Using end sequence specific primer allowed the overlapping clones to be oriented with respect to their T7 and SP6 ends. This procedure was used to confirm the assembly of the seven preliminary contigs by identifying BAC clones whose ends overlap. Moreover, this technique was used to incorporate additional BAC clones into existing contigs and linked some of the preliminary contigs together into larger contigs.

During the end sequence specific PCR stage of analysis we obtained seven CHORI-241 BAC clones (135 M23, 288 J19, 324 F02, 407 K07, 439 E06, 440 J07, and 455 C07) from our collaborators at Cornell University that were incorporated into our contig assembly and analysis. The overlaps of these clones were confirmed by overgo hybridization of end sequence specific overgo probes designed by our collaborators.

Table 3 End sequence specific PCR primers used to confirm contig assembly.

Primer Set	Accession number	Forward Primer	Reverse Primer	Annealing Temperature and conditions
024 A13 T7	CC875246	AGTGAATGAGAAGACAAGCCA	GGCAGGTTTTTGTGTGGA	TD 58-55°C, 54°C; 30 cy
102 G05 T7	CC875234	CTGCCACAACCCAAATGAG	AGCTTCCGTGCTCCTCTAGT	TD 58-55°C, 54°C; 30 cy
102 G05 SP6	CC875251	CTAGGTGAGTAGCGTATGCGG	TGTCCCAAGACCAAAAGTGTA	TD 58-55°C, 54°C; 30 cy
147 K21 T7	CC875220	GAACCAATCAGAAATGTTGGAG	AGTTCTCGCTCCTCCTCTATCT	TD 58-55°C, 54°C; 30 cy
147 K21 SP6	CC875240	AGCCTCAGTTCTGGGAAAAG	CAGGGTATTGCAGAGAGCAG	TD 58-57°C, 56°C; 30 cy
163 J11 T7	CC875238	GTCACATTTGAAAGCACGAAA	CCGACCCCATGTCATAAAA	TD 58-55°C, 54°C; 30 cy
163 J11 SP6	CC875237	CAATGGAAAAGCTCCCACTC	CACTGGCGGGATGGATTAT	TD 58-55°C, 54°C; 30 cy
288 J19 T7		CATAAGCATTTGGAAGCCAT	GTAAATCCGGCAAGAAGACA	50°C; 30 cy
288 J19 SP6		AGGTCTAGCACACTGCACTC	ACATTCAAGGAACGGCACTA	50°C; 30 cy
355 H03 T7	CC875236	ACCTCATTCGTGAAACTCTCC	CACCTCTGGGAACAGCTTATT	50°C; 30 cy
355 H03 SP6	CC875235	TGTTTGCTAAATGAATGTGGA	CTCTCTGATACTGTGAAAACTGG	TD 58-55°C, 54°C; 30 cy
359 L18 T7	CC875239	CAGAGTTCCAAAGAGCCAGG	TAGCACAGACGCATCGCAG	TD 58-55°C, 54°C; 30 cy
372 F10 T7	CC875224	TCAAGAGGCTTACAATCTGAGAAA	AGTCCCAAAACTTGCATTCC	TD 58-55°C, 54°C; 30 cy
372 F10 SP6	CC875222	ATCTAGCCACGGGCCACTT	AACCCAGGTGTCCAAGTCAG	TD 58-55°C, 54°C; 30 cy
382 H22 T7	CC875233	GAGACTAGAATGAGGATAAAGCTG	AGCTAAACCATGTTCTTCTCC	TD 58-55°C, 54°C; 30 cy
382 H22 SP6	CC875232	TCACTCTATCCCCACCATCTG	TTGTGCAAAGCCAAGATCC	TD 58-55°C, 54°C; 30 cy
394 D17 T7	CC875244	CGATACTACGGAGGACAGAAT	GCATTTCCTAATTGAGTTGG	TD 58-55°C, 54°C; 30 cy
394 D17 SP6	CC875243	TGAAGTGGATGGTGGAGTCA	GGGTGATGATGCCTTTCTCA	TD 58-55°C, 54°C; 30 cy
431 P04 T7	CC875228	ACCGGCCTGGGCTCTCAT	GCGCTACCAGGATCTCTCACTA	TD 58-55°C, 54°C; 30 cy
431 P04 SP6	CC875250	TGGGAGTGGAAGTGTATTTCA	GTCACTAGACCACAAACCCCT	TD 58-55°C, 54°C; 30 cy
464 F03 T7	CC875242	GCGATTGTTGGTCTGCTGTA	TCCATAAGTGTGGCAAGCTG	TD 58-55°C, 54°C; 30 cy
464 F03 SP6	CC875247	TACTGCCCACTCGGACTTCT	TATCTGCGCCTCAAGGAGAT	TD 58-55°C, 54°C; 30 cy
475 L06 T7	CC875230	ACAGCACAGGATACAGCACATC	TATATTAGGTTGGGGGTGGTCA	TD 58-55°C, 54°C; 30 cy
475 L06 SP6	CC875227	CATGAGTGTCCGATTAGGCTTAC	CATTCTTACCAGAGGAGGATGG	TD 58-55°C, 54°C; 30 cy
497 G11 T7	CC875223	AGAAAGGCACCTGTGGTCAG	GAATTGAGGCAGGAATTTGG	TD 58-55°C, 54°C; 30 cy
497 G11 SP6	CC875221	TGCTGGCTCTGAAGAAACAA	GCACTGATGGCTCCTGACTT	TD 58-55°C, 54°C; 30 cy
499 L11 T7	CC875231	ACAGTAAAATAGGGTGGAGGC	TGGCCTTGATTCCTCTCTGTA	TD 58-55°C, 54°C; 30 cy

Table 3 Continued

Primer Set	Accession number	Forward Primer	Reverse Primer	Annealing Temperature and conditions
499 L11 SP6	CC875229	CCTGACTTCTGCTGTTGACTC	GGAACCTCAATCTTGTCTGGT	TD 58-55°C, 54°C; 30 cy
528 E24 T7	CC875252	CAAAACAGACCCAAGCCATT	CAGTGGAAAGGGAGAGGGTA	TD 58-55°C, 54°C; 30 cy
528 E24 SP6	CC875226	ATCCCCAAGGACCAATCC	GCACAGAGACAAGATGACCCTA	TD 60-51°C, 50°C; 30 cy

Cycles is abbreviated as cy in this table.

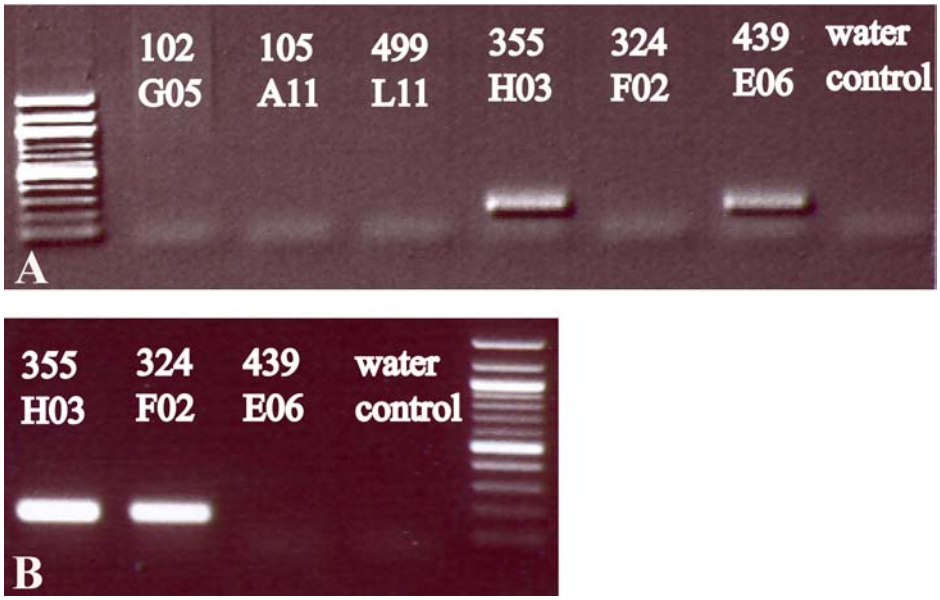

Fig. 5 End sequence specific PCR for BAC clone 355 H03. **A** 355 H03 SP6 end sequence specific PCR. Amplification of 355 H03 is a positive control and amplification of 439 E06 confirms the overlap of the SP6 end of 355 H03 with BAC clone 439 E06. **B** 355 H03 T7 end sequence specific PCR. Amplification of 355 H03 is a positive control and amplification of 324 F02 confirms the overlap of the T7 end of 355 H03 with BAC clone 324 F02.

Ultimately, the BAC clones were assembled into two contigs that span ELA region (Fig. 6). One contig spans the class II region and the minimum tiling path is comprised of nine BAC clones. The other contig spans the class I and class III region and the minimum tiling path contains 14 BAC clones. A single gap between the class II and class III regions prevented the joining of these two contigs (Fig. 7).

Identification of additional genes

Our collaborators at Cornell University analyzed the BAC clones which comprised the minimum tiling paths of the two contigs for the ELA region in order to identify additional genes within this region. Overgo probes were designed for 14 genes (five class I, five class II, and four class III) that exist within the MHC of other mammalian species and one equine microsatellite (Table 4). Overgo hybridization to secondary filters containing the BAC clones which represent the minimum tiling path of the two contigs resulted in the identification of all 14 genes and the one equine class II microsatellite. All of the genes were found to exist in the same region as predicted by the human genomic sequence of HLA (Figure 7). Additionally, *NOTCH4* and *GABBR1* were identified within BAC clones by PCR amplification (Figure 7). The addition of these genes increases the utility of the ELA BAC contig by providing more genes which can be used for comparative analysis of the organization and evolution of the MHC in mammals.

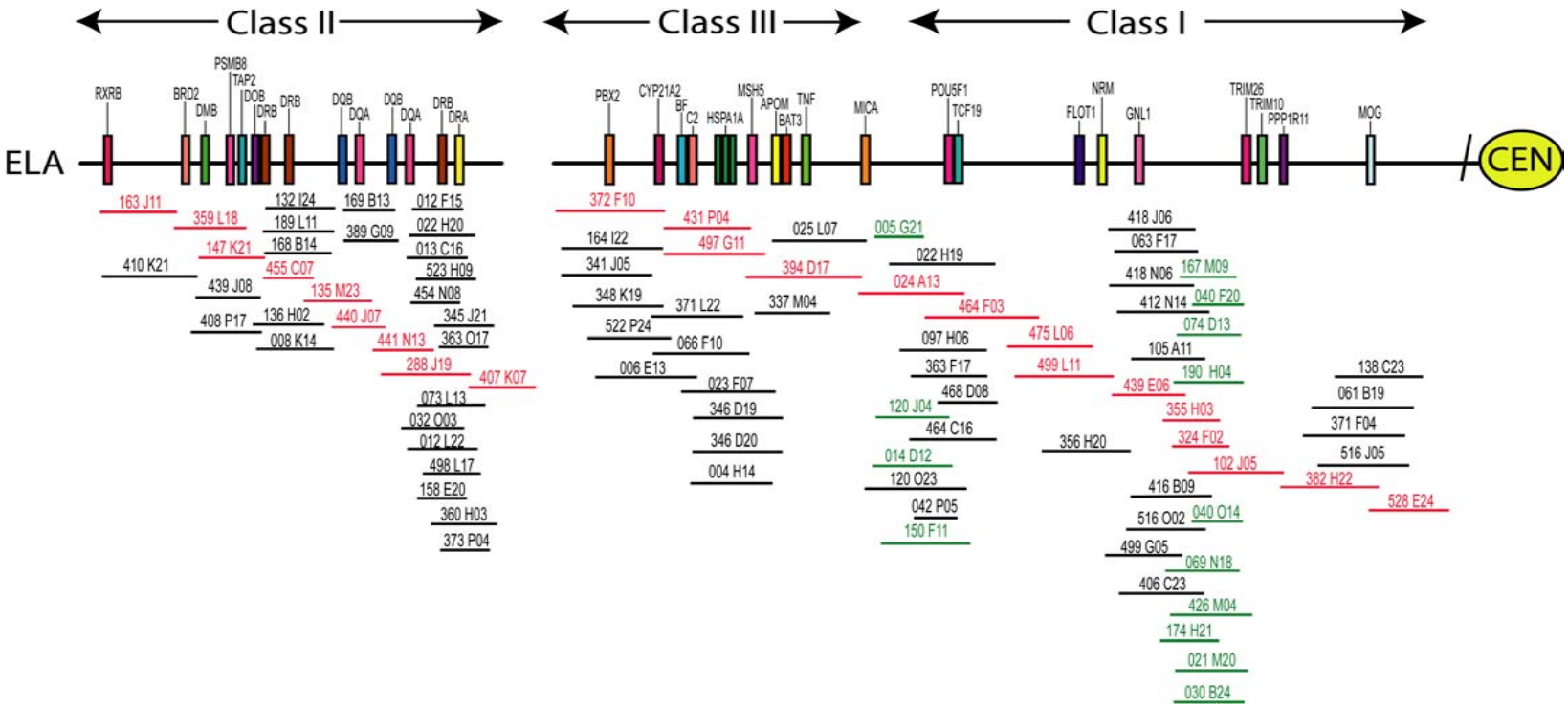

Fig. 6 Two BAC contigs that span the ELA region. The genes identified in this study are listed above the contigs. The BACs are depicted in their approximate location within the contigs. BACs in red represent the minimum tiling path for each contig. BACs in green contain class I genes and have been preliminarily positioned in the contig. This figure is not drawn to scale.

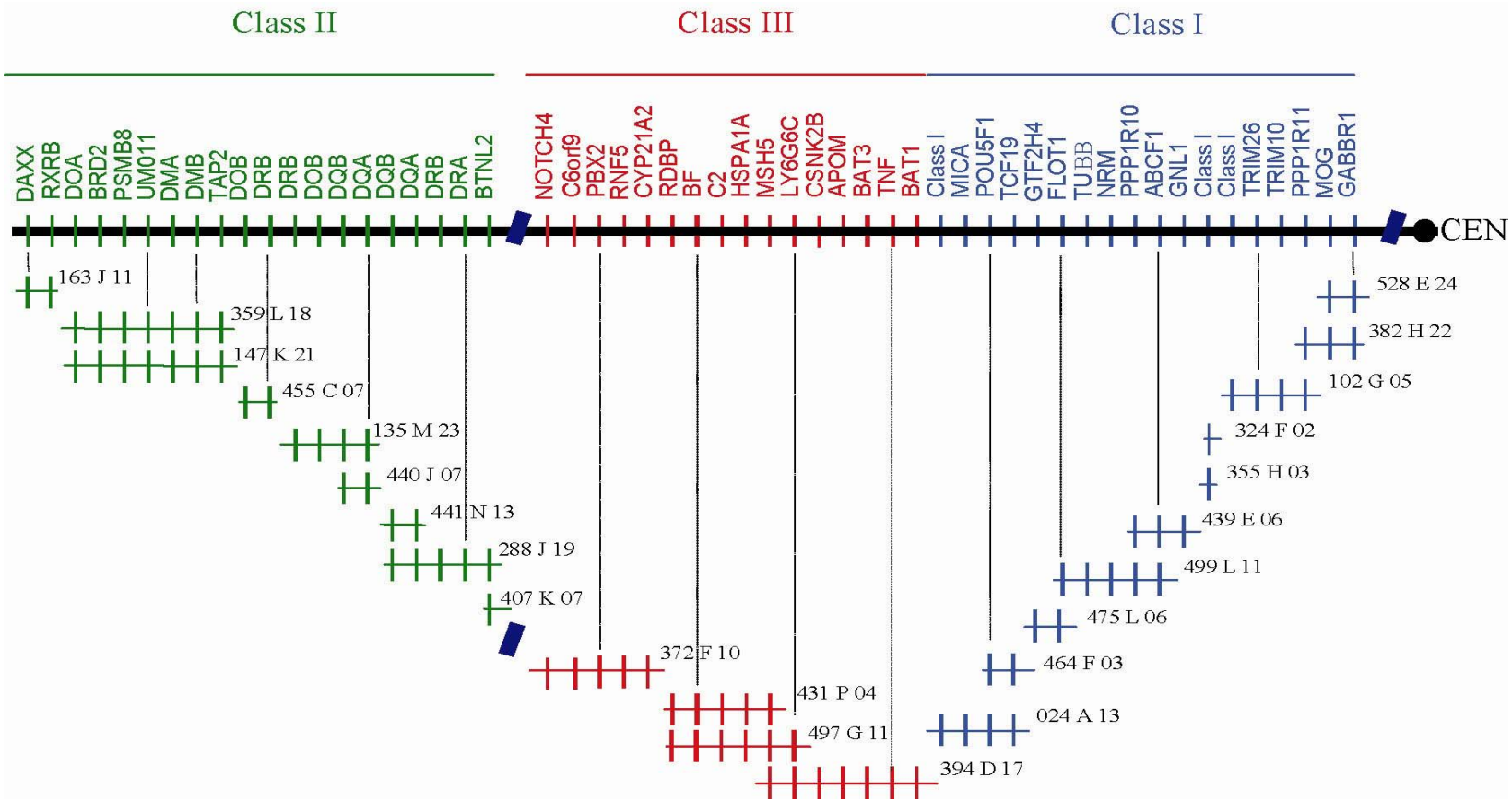

Fig. 7 Minimum tiling path of BAC contigs and gene map of ELA. Genes are shown in sequential order, as determined by overgo hybridization and end sequencing. The class II genes are green, the class III genes are red and the class I genes are blue. Genes are spaced equally because distance between genes has not been determined. The blue block indicates a gap.

Table 4 Overgo probes used to identify additional genes within the ELA BAC contig.

Gene	Gene Product	Accession No. (Species)	Forward Overgo Primer	Reverse Overgo Primer
ABCF1	ATP-binding cassette, sub-family F, member 1	10947134 (hs), 28499835 (mm), 10863746 (rn)	1) TTTGACCTTGAGATGCAGAATCGG 2) GCCTCATCACAGAAACCAACTGCC	1) AGAACTTCTGTGTGGGCCGATTCT 2) TCCACCACCCACACCTGGCAGTTG
BAT1	HLA-B associated transcript 1	19913439 (hs), 9790068 (mm), 6016839 (rn), 509402 (ss)	1) TGGCAGAGAACGATGTGGACAATG 2) GAAGCAGGTCATGATGTTCAGTGC	1) TCATAGTCCAAGAGCTCATTGTCC 2) CTCTTTGCTCAAGGTAGCACTGAA
BF	B-factor, properdin	14550403 (hs), 975231 (ss)	1) CTTTATCTTGGGCCTCTTGTCTGG 2) TGTCCTTCTGGCTTCTACCCGTAC	1) AGTCATGCTCACACCTCCAGACAA 2) TACGAATCTGCACAGGGTACGGGT
BTNL2	Butyrophilin-like 2	9624968 (hs)	1) CAGGCTACAATCTGTCTGGTGCAG 2) GCATCGCATCCAAGATAAAGATGG	1) AATAGGAAGGAGGCGACTGCACCA 2) TTCCGCATAGAACAGGCCATCTTT
CSNK2B	Casein kinase 2, beta polypeptide	26787971 (hs), 7106276 (mm), 13591929 (rn)	1) TCGGCTTTTTGCGCTGTAGTGGTC 2) GGTGAAACTCTACTGCCCCAAGTG	1) TCCAAGGAACCGCAGAGACCACTA 2) TGTGTACACGTCCATGCACTTGGG
DAXX	Death-associated protein 6	11908170 (hs), 6681134 (mm), 18148938 (rn)	1) GAAGAGTTCCTTGAACTGTGTAAG 2) CACTGTATGTGGCAGAGATTCGGC	1) GGTCTGATGTCTGCATCTTACACA 2) TCCTTTTCCTGCAGCCGCCGAATC
DMA	MHC, class II, DM alpha	18765714 (hs), 6754119 (mm)	TCACGCTGAAGCCCCTGGAGTTTG	AAAGTGTTGGGCTTGCCAAACTCC
DOA	MHC class II, DO alpha	4504400 (hs), 6680150 (mm)	CCCCCTGTGATCAATATCACCTGG	TTTGGCCGTTGCGCAGCCAGGTGA
GPSM3	G-protein signalling modulator 3	31389770 (ec)	TCCTGGAGTTGCTGCTGAGAGTTC	ATTCGACCCCCACCATGAACTCTC

Table 4 Continued

Gene	Gene Product	Accession No. (Species)	Forward Overgo Primer	Reverse Overgo Primer
GTF2H4	General transcription factor II H, polypeptide 4	27498326 (hs), 6754093 (mm)	1) GAACCGAGTACACCTACAATGCAG 2) GGTGTCCTGTATAACCAGTTCCTG	1) GAACTCCTGCAGATTCCTGCATTG 2) CAAAGTCCACTTGCGACAGGAACT
MICA	MHC class I-related chain A	18033157 bt), 6624722 (ss)	1) TTTTGCTGAGGGACACTTGGATGG 2) AGTCTGGGGATGTCCTGCCTGATG	1) GCGCAGGAAGGCCTGACCATCCAA 2) TGGTAGGTCCCATTCCCATCAGGC
PPP1R11	Protein phosphatase 1, regulatory subunit 11	11386174 (hs), 18390326 (mm)	1) CGGACGCTTACTATGAAACTTCGG 2) AGGATGAAGAAGAGGGCTGTGGTC	1) TCTCTGGCTTCCGTTTCCGAAGTT 2) CGTACACAGTGCGTATGACCACAG
RNF5	Ring finger protein 5	13026452 (hs), 9507058 (mm)	1) TCAGTGTGTGTGGCCACCTGTACT 2) GCCAGAAGCCCCAGGATCCCAGAT	1) TGAAGACATGGCCAACAGTACAGG 2) CGGGGTGGAGTTTTTAATCTGGGA
TAP2	Transporter 2, ATP-binding cassette	16610227 (hs), 19171651 (bt)	1) AGGCTTCTTCACCTTCACCATGTC 2) CATCCTGGATGAGGCTACTAGTGC	1) CCGCAAGTTGATTCGAGACATGGT 2) ACACTCCACATCCAGGGCACTAGT
TUBB	Tubulin beta	Unpublished seq (ec)	ATCCAGGAGCTGTTCAAGCGCATC	CTGTGAATTGCTCCGAGATGCGCT
UM011	Microsatellite	6180135 (ec)	TTGGTGAGGATTAGGGGTTTTCCC	CAGTTGGGAGGGAAATGGGAAAAC

In cases where there is more than one overgo primer pair per gene the overgo primer pairs are denoted by numbers to indicate the primers that form a pair. Species were abbreviated as followed: human (hs), mouse (mm), rat (rn), pig (ss), cattle (bt), and horse (ec).

Characterization of the contigs and development of a physical map for ELA

The minimum tiling path of the class II contig is comprised of nine overlapping BAC clones. Restriction fragment length analysis of the BAC clones within the class II contig gave an estimate of approximately 800 kb of DNA in the class II contig (Fig. 1). One equine microsatellite and 19 genes (13 class II and 6 non-class II genes) were mapped to the class II region. The order of the genes in this region is presented in Figure 7. Two class I antigen processing and presentation genes, *PSMB8* and *TAP2*, were identified in the class II region of ELA. These results follow the same trend as all other mammalian MHCs investigated to date. Southern blot analysis indicated the presence of at least one *DRA* locus, two *DQA* and *DQB* loci, and three *DRB* loci. However, the exact locations of these loci and whether they were functional genes or pseudogenes could not be determined in this study.

The class III region was estimated to contain approximately 430 kb of DNA as deduced by restriction fragment analysis of the four BAC clones spanning the minimum tiling path of the class III portion of the second contig (Fig. 2). Sixteen genes were mapped to the class III region and end sequence analysis confirmed the presence of the following six genes: *C2*, *LY6G6C*, *APOM*, *RDBP*, *MSH5*, and *CYP21A2*. The gene order identified within the class III region is similar to the order found in other mammalian species (Figure 7). In particular the gene order of the ELA class III region appears to be highly conserved with that of HLA. Southern blot hybridizations of the class III BAC clones with *C4* and *CYP21A2* overgo probe revealed no evidence of a

duplication of the *C4-CYP21A2* region, which has been observed in the class III region of other mammalian species (Fig. 4A and 4B).

BAC clones encompassing the MHC class I region contained about 1.2 Mb of DNA, extending from *MICA* to *GABBR1*. Fifteen anchor genes were identified in this region, three of which were confirmed by end sequences (*NRM*, *PPP1R10*, and *MOG*). The specific gene order for the class I region is presented in Fig. 7. Three regions of the second contig contained class I sequences. The regions located near *MICA* and *GNL1* each appear to contain multiple class I genes. A third region, located near *TRIM26*, appears to contain a single class I gene. Characterization of the class I genes contained within these regions has been initiated to determine how the distribution of ELA class I genes relates to the distribution of class I sequences in other species.

FISH

FISH was used to anchor each BAC contig on ECA20 and to identify and orient the different MHC regions relative to the centromere. FISH was performed using 14 BAC clones (three class I, nine class II, and two class III) strategically selected from different regions of the two BAC contigs. All BAC clones mapped to ECA20q21 confirming the previous localization of the horse MHC by in situ hybridization (Ansari et al. 1988; Makinen et al. 1989). Dual color FISH using representative clones from the class I and class II regions showed the class I region to be proximal to the class II region (Fig. 8a). To confirm the location of the class III region, dual color interphase FISH was

performed. The results indicated that the class III region is located between the class I and class II regions of the ELA region (Fig. 8b).

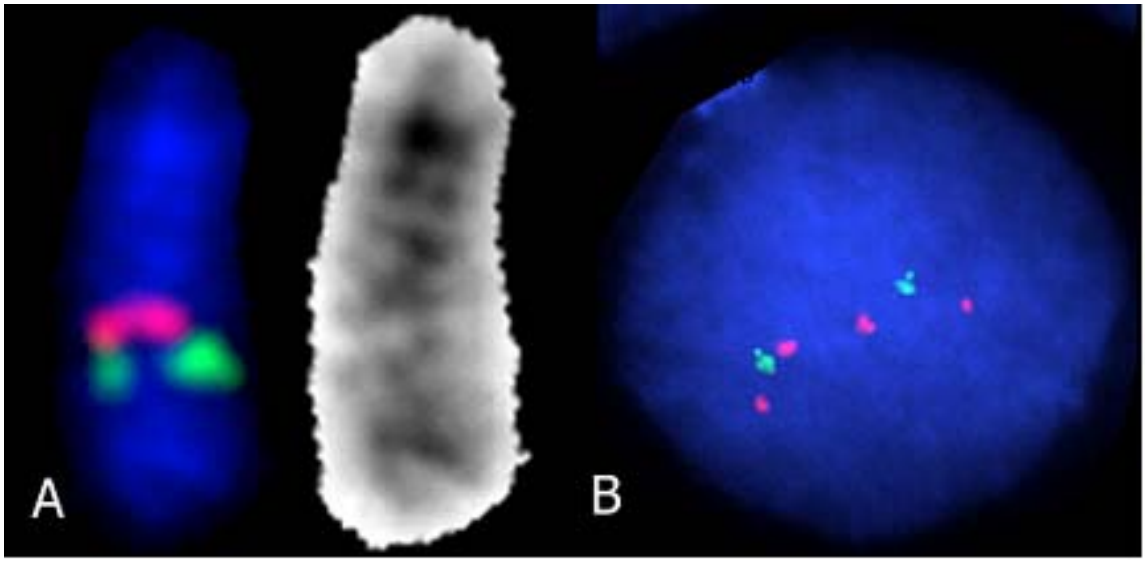

Fig. 8 Localization of horse MHC BAC clones to ECA20q to demonstrate the chromosomal orientation of the horse MHC. **A** Dual color metaphase FISH of fluorescently-labeled DNA from BAC clones showing the chromosomal orientation of the horse MHC on ECA20q21. Class I MHC BAC 528 E24 shows red signal and the class II MHC BAC 147 K21 a green signal. **B** Dual color FISH to equine interphase nuclei. This technique is demonstrating that the equine MHC class III region (BAC 431 P04 labeled green) is located between the class I (BAC 528 E24 labeled red) and II (BAC147 K21 labeled red) regions.

Discussion

The two BAC contigs described herein provide the first comprehensive physical map of the horse MHC. The class II region contig spans approximately 800 kb, while the class III – class I contig spans approximately 1.6 Mb. The gap between the ELA class II and class III regions is estimated to be approximately 170 kb based on the distance between *BTNL2* (distal marker of class II region) and *NOTCH4* (proximal marker of class III region) in the human MHC. Assuming this estimate to be correct, the overall size of the ELA region is approximated to be 2.6 Mb of DNA, a figure smaller than the 3.6 Mb estimated for the human MHC (MHC Sequencing Consortium 1999), but slightly larger than the pig MHC (Chardon et al. 1999a). The size difference between ELA and other mammalian MHC regions is likely to be the result of a reduction in the amount of sequence that exists between genes because based upon comparative analyses it does not appear that ELA has lost any genes within the MHC region.

Studies of domesticated species including cow (Lewin et al. 1999) and pig (Chardon et al. 1999a) have reported multiple loci for many of the D region genes including *DQA*, *DQB*, and *DRB*, and the horse appears to share this phenomenon. Southern blot analysis of ELA class II BAC clones and independent assignment of class II genes to non-overlapping BAC clones by filter hybridization confirmed the existence of at least two *ELA-DQB* loci (Fig. 3), described earlier by Horin and Matiasovic (2002). The same techniques provided evidence for the presence of at least one locus for *ELA-DRA* and *ELA-DQA*, supporting previously published results (Albright-Fraser et al. 1996; Fraser

and Bailey 1996; Fraser and Bailey 1998). Additionally, at least three *ELA-DRB* loci were identified, which is in agreement with previous studies of other domesticated equids (Fraser and Bailey 1996). However, recent studies found that Andalusian stallions and Przewalski's horses appear to have at most two *DRB* genes (Hedrick et al. 1999). Thus, it is possible that the number of *DRB* loci varies between equine MHC haplotypes, as has been found in other species such as cattle (Ellis and Ballingall 1999; Gongora et al. 1997). Additionally, the class I antigen processing and presentation genes, *TAP2* and *PSMB8*, were identified within the class II region which agrees with the findings in all other mammalian species studied to date. These results strengthen the idea that the close linkage of the class I processing and presentation genes (*TAP1*, *TAP2*, *PSMB8*, and *PSMB9*) with the class II region in the mammalian MHC is likely a derived organization that occurred after the divergence of birds. Overall, our results corroborate and extend the findings of previously published characterizations of equine MHC class II genes.

Fingerprinting analysis indicates that the size of the class III region in the horse is smaller relative to that of human and mouse (Amadou et al. 1999; MHC Sequencing Consortium 1999). This reduction in size is most likely the result of less sequence existing between the genes. Since, there does not appear to be any genes missing from the class III region of ELA. Overall, the gene order within the ELA class III region appears to be well conserved with class III regions of human, mouse, and pig, which is consistent with the suggestion that this region of the mammalian genome is

evolutionarily more conserved than the other regions of the MHC (Peelman et al. 1996). This conservation is suggestive of a potential functional significance related to the organization of these genes within the MHC region. Southern blot hybridization of probes for class III genes did not discover any evidence of duplicated genes in the class III region. In particular the *C4* and *CYP21A2* genes which are tandemly duplicated in certain human and mouse haplotypes were investigated. However, no evidence of *C4* or *CYP21A2* duplication was discovered, which is similar to what is observed in the pig class III region (Peelman et al. 1996).

Of the 15 genes localized to the equine MHC class I region, seven genes (*POU5F1*, *TCF19*, *GNL1*, *TRIM26*, *TRIM10*, *PPP1R11*, and *MOG*) were previously identified as highly conserved framework genes in human and mouse (Amadou 1999). The framework hypothesis predicts that functionally important genes in the class I region form a scaffold of framework genes and within the framework scaffold are regions referred to as duplication blocks, where class I genes have been inserted, duplicated, and expanded. Three major duplication blocks designated alpha (between *PPP1R11* and *MOG*), beta (between *BAT1* and *POU5F1*), and kappa (between *GNL1* and *TRIM26*) have been identified in a number of species. However, the number of class I genes found within these regions varies greatly between species (Kulski et al. 2002). In this study, the ELA regions shown to contain class I genes coincide with the beta and kappa blocks, but no clear evidence of class I genes in the alpha block was found. This arrangement is similar to that reported in pigs where class I genes are only located in the

beta and kappa block (Velten et al. 1999). Detailed analysis of the ELA class I region will be necessary to confirm the framework organization suggested in this study, but the ELA data to date appear to support the conservation of framework genes and duplication blocks across species and the idea that the genomic organization of the class I region has functional and evolutionary significance. Additionally, the results of this study provide an opportunity to investigate the potential reasons why certain genomes such as the pig and the horse, do not utilize all the available duplication blocks for the expansion of class I genes.

The gene order and organizational features of the horse MHC described here are in general agreement with that described for the primates (MHC Sequencing Consortium 1999; Leelayuwat et al. 1993) and carnivores (Beck et al. 2001; Wagner 2003), although ELA appears to be reduced in size. This observation implies that the disruptions of the MHCs seen in pigs and ruminants occurred after divergence of the Artiodactyls in the mammalian lineage. In swine the MHC is disrupted by a centromere, and in ruminants, by a large chromosomal inversion (Band et al. 1998; Skow et al. 1996; Smith et al. 1995). Interestingly, characterization of the chicken MHC also shows a disruption that is not observed in the passerine birds (Kaufman et al. 1999b; Shiina et al. 2004b). The genomic structure of the horse, human, rat, and mouse class II and III regions seems to be highly conserved (Amadou et al. 1999; MHC Sequencing Consortium 1999; Hurt et al. 2004). These observations suggest that this region may have existed in a primordial

ancestor in a similar organization and was not subjected to major rearrangement during evolution in these species.

In conclusion, comparative analysis of the equine MHC region has demonstrated significant conservation of gene order and genomic structure relative to other mammalian species, except the Artiodactyla. The construction of two BAC contigs has increased our knowledge of the gene content and organization of the MHC region of the horse and provided the sequence-ready templates required for detailed analysis of this important region of the genome. We anticipate that further characterization of ELA will provide valuable information on the functional genomics and evolution of the MHC of the horse and contribute to our understanding of the relationship between the genes of the equine MHC and the equine immune response.

CHAPTER III

CONCLUSIONS

The MHC is one of the most well studied regions of the vertebrate genome due to its gene dense nature, role in immune function, and association with disease resistance and susceptibility. Gene mapping and genomic sequencing of the MHC region in several divergent species has facilitated comparative analyses focusing on the overall conservation and stability of MHC organization as well as gene content. It is through these studies that the evolution of this region can be accurately assessed. Nonetheless, future investigations involving additional vertebrate species are needed to accurately reconstruct the evolutionary history of the MHC region as a whole and the class I and class II genes in particular. To date the MHC has not been extensively investigated in any species of Perissodactyla. Therefore, we provide the first and only analysis and characterization of the MHC in *Equus caballus*, a species of Perissodactyla.

Two BAC contigs that span the equine MHC were developed in this study. These BAC contigs provide a sequence-ready template for complete genomic sequencing of this region. The availability of genomic sequence for this region would dramatically aid in the further characterization of ELA. Also, it would provide the materials necessary to develop MHC-specific molecular markers (gene-specific and microsatellite) that could be used to investigate potential associations between the MHC and natural disease resistance. Additionally, the availability of genomic sequence would tremendously aid targeted candidate gene studies that could potentially define MHC

disease association to a specific gene, haplotype, or allele. While complete genomic sequencing of ELA would provide the opportunity to do all the aforementioned experiments, targeted sequencing of specific BAC clones within these contigs will also allow for further investigations of specific genes or regions of the MHC.

While developing the BAC contigs it was necessary to identify some of the genes contained within the BAC clones to confirm their genomic location and to assemble them into contigs. Thus, the first physical map of ELA was developed in this study. The ELA physical map that was generated identified ~20% of the genes that exist in HLA and provided an excellent opportunity for comparative analysis of the human and horse MHC regions. Comparative analysis with HLA revealed a striking conservation of gross genomic organization, gene content, and gene order. The human and horse MHC regions appear to be well conserved with respect to gene order. All of the genes investigated in our study were found in the same relative position that they exist in HLA. In fact, while assembling the class III contig the gene order was so well conserved that overlapping BAC clones could be predicted strictly based upon the gene content and order that exists in the class III region of HLA. Additional comparative analyses of the equine MHC with other mammalian MHCs demonstrated a considerable conservation of gene content; however the number of class I and class II genes did vary from species to species.

Prior to the current study, investigations primarily focused on the horse MHC class I and class II genes existed. In contrast, the primary objective of this study was to provide a comprehensive analysis of ELA, including mapping and annotating the overall

gene content. However, the duplication/deletion nature of the class I and class II genes make it difficult to precisely map these genes within each BAC clone. Therefore, sequencing of the BAC clones that contain class I and class II genes will be necessary for precise mapping of these genes with respect to the other ELA genes especially given that the exact number of class I genes in ELA has yet to be determined.

In conclusion, comparative analyses of the equine MHC region have demonstrated significant conservation of gene content and gene order, relative to other mammalian species. However, the genomic architecture of ELA appears to most similar to humans, where the MHC exists as a single contiguous segment and lacks disruptions seen in ruminants, pig, and cats. The assembly of two BAC contigs and the development of a physical map for ELA has significantly increased our knowledge of the organization and content of the MHC region of the horse. Additionally, the two BAC contigs provide the templates for complete genomic sequencing of this important genomic region. Herein we have provided a thorough molecular framework for future studies involving additional characterization of ELA through both mapping and sequencing, thereby providing the opportunity for additional evolutionary analyses as well as association studies aimed at assessing the relationship between ELA and disease.

REFERENCES

Albright-Fraser DG, Reid R, Gerber V, Bailey E (1996) Polymorphism of DRA among equids. Immunogenetics 43:315-317

Alexander AJ, Bailey E, Woodward JG (1987) Analysis of the equine lymphocyte antigen system by Southern blot hybridization. Immunogenetics 25:47-54

Alfonso C, Karlsson L (2000) Nonclassical MHC class II molecules. Annu Rev Immunol 18:113-142

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Amadou C (1999) Evolution of the Mhc class I region: the framework hypothesis. Immunogenetics 49:362-367

Amadou C, Kumanovics A, Jones EP, Lambracht-Washington D, Yoshino M, Lindahl KF (1999) The mouse major histocompatibility complex: some assembly required. Immunol Rev 167:211-221

Andersson L, Lunden A, Sigurdardottir S, Davies CJ, Rask L (1988) Linkage relationships in the bovine MHC region. High recombination frequency between class II subregions. Immunogenetics 27:273-280

Ansari HA, Hediger R, Fries R, Stranzinger G (1988) Chromosomal localization of the major histocompatibility complex of the horse (ELA) by in situ hybridization. Immunogenetics 28:362-364

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. Crit Rev Immunol 17:179-224

Bailey E, Marti E, Fraser DG, Antczak DF, Lazary S (2000) Immunogenetics of the horse. In: Bowling AT, Ruvinsky A (eds) The Genetics of the Horse. CAB International, Wallingford, UK, pp 123-155

Ballingall K, MacHugh N, Taracha E, Mertens B, McKeever D (2001) Transcription of the unique ruminant class II major histocompatibility complex-DYA and DIB genes in dendritic cells. Eur J Immunol 31:82-86

Ballingall KT, Ellis SA, MacHugh ND, Archibald SD, McKeever DJ (2004) The DY genes of the cattle MHC: expression and comparative analysis of an unusual class II MHC gene pair. Immunogenetics 55:748-755

Band M, Larson JH, Womack JE, Lewin HA (1998) A radiation hybrid map of BTA23: identification of a chromosomal rearrangement leading to separation of the cattle MHC class II subregions. Genomics 53:269-275

Barbis DP, Maher JK, Stanek J, Klaunberg BA, Antczak DF (1994) Horse cDNA clones encoding two MHC class I genes. Immunogenetics 40:163

Beck S, Trowsdale J (1999) Sequence organisation of the class II region of the human MHC. Immunol Rev 167:201-210

Beck TW, Menninger J, Voigt G, Newmann K, Nishigaki Y, Nash WG, Stephens RM, Wang Y, de Jong PJ, O'Brien SJ, Yuhki N (2001) Comparative feline genomics: a BAC/PAC contig map of the major histocompatibility complex class II region. Genomics 71:282-295

Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, Gates M, Postlethwait JH, Klein J (1997) Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, Danio rerio. Immunogenetics 46:129-134

Blanchong CA, Chung EK, Rupert KL, Yang Y, Yang Z, Zhou B, Moulds JM, Yu CY (2001) Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologues, Slp and C4. Int Immunopharmacol 1:365-392

Bover KH (1927) Homoisotransplantation van Epidermis bei eineigen Zwillingen. Beitr. Klin. Chir 141:442-447

Brocke P, Garbi N, Momburg F, Hammerling GJ (2002) HLA-DM, HLA-DO and tapasin: functional similarities and differences. Curr Opin Immunol 14:22-29

Brostrom H, Fahlbrink E, Dubath ML, Lazary S (1988) Association between equine leucocyte antigens (ELA) and equine sarcoid tumors in the population of Swedish halfbreds and some of their families. Vet Immunol Immunopathol 19:215-223

Burch GH, Gong Y, Liu W, Dettman RW, Curry CJ, Smith L, Miller WL, Bristow J (1997) Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. Nat Genet 17:104-108

Busch R, Doebele RC, Patil NS, Pashine A, Mellins ED (2000) Accessory molecules for MHC class II peptide loading. Curr Opin Immunol 12:99-106

Carpenter S, Baker JM, Bacon SJ, Hopman T, Maher J, Ellis SA, Antczak DF (2001) Molecular and functional characterization of genes encoding horse MHC class I antigens. Immunogenetics 53:802-809

Chambers G, Ellsmore VA, O'Brien PM, Reid SW, Love S, Campo MS, Nasir L (2003) Association of bovine papillomavirus with the equine sarcoid. J Gen Virol 84:1055-1062

Chardon P, Renard C, Gaillard CR, Vaiman M (2000) The porcine major histocompatibility complex and related paralogous regions: a review. Genet Sel Evol 32:109-128

Chardon P, Renard C, Vaiman M (1999a) The major histocompatibility complex in swine. Immunol Rev 167:179-192

Chardon P, Rogel-Gaillard C, Cattolico L, Duprat S, Vaiman M, Renard C (2001) Sequence of the swine major histocompatibility complex region containing all non-classical class I genes. Tissue Antigens 57:55-65

Chardon P, Rogel-Gaillard C, Peelman L, Yerle M, Velten FW, Renard C, Vaiman M (1999b) Physical organization of the pig major histocompatibility complex class II region. Immunogenetics 50:344-348

Choonhakarn C, Romphruk A, Puapairoj C, Jirarattanapochai K, Leelayuwat C (2002) Haplotype associations of the major histocompatibility complex with psoriasis in Northeastern Thais. Int J Dermatol 41:330-334

Chowdhary BP, Raudsepp T, Kata SR, Goh G, Millon LV, Allan V, Piumi F, Guerin G, Swinburne J, Binns M, Lear TL, Mickelson J, Murray J, Antczak DF, Womack JE, Skow LC (2003) The first-generation whole-genome radiation hybrid map in the horse identifies conserved segments in human and mouse genomes. Genome Res 13:742-751

Clark MS, Shaw L, Kelly A, Snell P, Elgar G (2001) Characterization of the MHC class I region of the Japanese pufferfish (*Fugu rubripes*). Immunogenetics 52:174-185

Collier S, Sinnott PJ, Dyer PA, Price DA, Harris R, Strachan T (1989) Pulsed field gel electrophoresis identifies a high degree of variability in the number of tandem 21-hydroxylase and complement C4 gene repeats in 21-hydroxylase deficiency haplotypes. Embo J 8:1393-1402

Crepaldi T, Crump A, Newman M, Ferrone S, Antczak DF (1986) Equine T lymphocytes express MHC class II antigens. J Immunogenet 13:349-360

Curik I, Fraser D, Eder C, Achmann R, Swinburne J, Crameri R, Brem G, Solkner J, Marti E (2003) Association between the MHC gene region and variation of serum IgE levels against specific mould allergens in the horse. Genet Sel Evol 35 Suppl 1:S177-190

Dausset J (1958) Iso-leuco-anticorps. Acta. Haematol. 20:156-166

Dawkins R, Leelayuwat C, Gaudieri S, Tay G, Hui J, Cattley S, Martinez P, Kulski J (1999) Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunol Rev 167:275-304

Dawkins RL, Kay PH, Christiansen FT (1983) Immunogenetics of rheumatoid arthritis. Ann Acad Med Singapore 12:155-163

Doherty PC, Zinkernagel RM (1975) Enhanced immune surveillance in mice heterozygous at the H-2 gene complex. Nature 256:50-52

Donaldson WL, Oriol JG, Plavin A, Antczak DF (1992) Developmental regulation of class I major histocompatibility complex antigen expression by equine trophoblastic cells. Differentiation 52:69-78

Dunham I, Sargent CA, Dawkins RL, Campbell RD (1989) Direct observation of the gene organization of the complement C4 and 21-hydroxylase loci by pulsed field gel electrophoresis. J Exp Med 169:1803-1818

Edwards SV, Wakeland EK, Potts WK (1995) Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. Proc Natl Acad Sci USA 92:12200-12204

Ellis SA, Ballingall KT (1999) Cattle MHC: evolution in action? Immunol Rev 167:159-168

Ellis SA, Holmes EC, Staines KA, Smith KB, Stear MJ, McKeever DJ, MacHugh ND, Morrison WI (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. Immunogenetics 50:319-328

Ellis SA, Martin AJ, Holmes EC, Morrison WI (1995) At least four MHC class I genes are transcribed in the horse: phylogenetic analysis suggests an unusual evolutionary history for the MHC in this species. Eur J Immunogenet 22:249-260

Flajnik MF (1998) Churchill and the immune system of ectothermic vertebrates. Immunol Rev 166:5-14

Flajnik MF, Kasahara M (2001) Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity 15:351-362

Flajnik MF, Ohta Y, Namikawa-Yamada C, Nonaka M (1999) Insight into the primordial MHC from studies in ectothermic vertebrates. Immunol Rev 167:59-67

Forbes SA, Trowsdale J (1999) The MHC quarterly report. Immunogenetics 50:152-159

Fraser DG, Bailey E (1996) Demonstration of three DRB loci in a domestic horse family. Immunogenetics 441-445

Fraser DG, Bailey E (1998) Polymorphism and multiple loci in the horse DQA gene. Immunogenetics 487-490

Goldberg AL, Rock KL (1992) Proteolysis, proteasomes and antigen presentation. Nature 357:375-379

Gongora R, Figueroa F, Klein J (1997) Complex origin of the HLA-DR10 haplotype. J Immunol 159:6044-6051

Goodrich L, Gerber H, Marti E, Antczak DF (1998) Equine sarcoids. Vet Clin North Am Equine Pract 14:607-623

Gorer PA (1937) The genetic and antigenic basis for tumor transplantation. J. Pathol. Bacteriol. 44:691-697

Gorer PA, Lyman S, Snell GD (1948) Studies on the genetic and antigenic basis of tumour transplantation: linkage between a histocompatibility gene and 'Fused' in mice. Proc R Soc London Ser B 135:499-505

Graser R, Vincek V, Takami K, Klein J (1998) Analysis of zebrafish Mhc using BAC clones. Immunogenetics 47:318-325

Grimsley C, Mather KA, Ober C (1998) HLA-H: a pseudogene with increased variation due to balancing selection at neighboring loci. Mol Biol Evol 15:1581-1588

Gruen JR, Weissman SM (2001) Human MHC class III and IV genes and disease associations. Front Biosci 6:D960-972

Gunther E, Walter L (2001) The major histocompatibility complex of the rat (*Rattus norvegicus*). Immunogenetics 53:520-542

Han CS, Sutherland RD, Jewett PB, Campbell ML, Meincke LJ, Tesmer JG, Mundt MO, Fawcett JJ, Kim U, Deaven LL, Doggett NA (2000) Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. Genome Research 10:741-721

Hedrick PW, Parker KM, Miller EL, Miller PS (1999) Major histocompatibility complex variation in the endangered Przewalski's horse. Genetics 152:1701-1710

Herberg JA, Beck S, Trowsdale J (1998) TAPASIN, DAXX, RGL2, HKE2 and four new genes (BING 1, 3 to 5) form a dense cluster at the centromeric end of the MHC. J Mol Biol 277:839-857

Hess M, Goldammer T, Gelhaus A, Ried K, Rappold G, Eggen A, Bishop MD, Schwerin M, Horstmann RD (1999) Physical assignment of the bovine MHC class IIa and class IIb genes. Cytogenet Cell Genet 85:244-247

Horin P, Matiasovic J (2002) A second locus and new alleles in the major histocompatibility complex class II (ELA-DQB) region in the horse. Anim Genet 33:196-200

Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC, Jr., Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S (2004) Gene map of the extended human MHC. Nat Rev Genet 5:889-899

Hughes AL, Hughes MK (1995) Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics 42:233-243

Hughes AL, Hughes MK, Howell CY, Nei M (1994) Natural selection at the class II major histocompatibility complex loci of mammals. Philos Trans R Soc Lond B Biol Sci 346:359-366; discussion 366-357

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167-170

Hughes AL, Nei M (1989a) Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol Biol Evol 6:559-579

Hughes AL, Nei M (1989b) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958-962

Hughes AL, Yeager M (1998a) Natural selection and the evolutionary history of major histocompatibility complex loci. Front Biosci 3:d509-516

Hughes AL, Yeager M (1998b) Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet 32:415-435

Hughes AL, Yeager M, Ten Elshof AE, Chorney MJ (1999) A new taxonomy of mammalian MHC class I molecules. Immunol Today 20:22-26

Hurt P, Walter L, Sudbrak R, Klages S, Muller I, Shiina T, Inoko H, Lehrach H, Gunther E, Reinhardt R, Himmelbauer H (2004) The genomic sequence and comparative analysis of the rat major histocompatibility complex. Genome Res 14:631-639

Ioannidu S, Walter L, Dressel R, Gunther E (2001) Physical map and expression profile of genes of the telomeric class I gene region of the rat MHC. J Immunol 166:3957-3965

Kasahara M, Suzuki T, Pasquier LD (2004) On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. Trends Immunol 25:105-111

Kaufman J, Jacob J, Shaw I, Walker B, Milne S, Beck S, Salomonsen J (1999a) Gene organisation determines evolution of function in the chicken MHC. Immunol Rev 167:101-117

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999b) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923-925

Kawaguchi H, Zaleska-Rutczynska Z, Figueroa F, O'HUigin C, Klein J (1992) C4 genes of the chimpanzee, gorilla, and orang-utan: evidence for extensive homogenization. Immunogenetics 35:16-23

Klein J (2001) George Snell's first foray into the unexplored territory of the major histocompatibility complex. Genetics 159:435-439

Klein J, O'hUigin C (1994) The conundrum of nonclassical major histocompatibility complex genes. Proc Natl Acad Sci USA 91:6251-6252

Klein J, Sato A (1998) Birth of the major histocompatibility complex. Scand J Immunol 47:199-209

Klein J, Sato A, O'hUigin C (1998) Evolution by gene duplication in the major histocompatibility complex. Cytogenet Cell Genet 80:123-127

Kulski JK, Gaudieri S, Dawkins RL (2000) Using alu J elements as molecular clocks to trace the evolutionary relationships between duplicated HLA class I genomic segments. J Mol Evol 50:510-519

Kulski JK, Martinez P, Longman-Jacobsen N, Wang W, Williamson J, Dawkins RL, Shiina T, Naruse T, Inoko H (2001) The association between HLA-A alleles and an Alu dimorphism near HLA-G. J Mol Evol 53:114-123

Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev 190:95-122

Kumanovics A, Madan A, Qin S, Rowen L, Hood L, Fischer Lindahl K (2002) QUOD ERAT FACIENDUM: sequence analysis of the H2-D and H2-Q regions of 129/SvJ mice. Immunogenetics 54:479-489

Kumanovics A, Takada T, Lindahl KF (2003) Genomic organization of the mammalian MHC. Annu Rev Immunol 21:629-657

Kuroda N, Figueroa F, O'HUigin C, Klein J (2002) Evidence that the separation of Mhc class II from class I loci in the zebrafish, Danio rerio, occurred by translocation. Immunogenetics 54:418-430

Kuroda N, Naruse K, Shima A, Nonaka M, Sasaki M (2000) Molecular cloning and linkage analysis of complement C3 and C4 genes of the Japanese medaka fish. Immunogenetics 51:117-128

Kurotaki T, Narayama K, Oyamada T, Yoshikawa H, Yoshikawa T (2000) The kinetics of Langerhans cells in equine insect hypersensitivity "Kasen". J Vet Med Sci 62:561-564

Kuwahara Y, Kitoh K, Kobayashi R, Iwata J, Ohne R, Hosokawa-Kanai T, Matsumoto Y, Kitagawa H, Sasaki Y (2000) Genotyping of feline MHC (FLA) class II DRB by PCR-RFLP method using group-specific primers. J Vet Med Sci 62:1283-1289

Leelayuwat C, Zhang WJ, Abraham LJ, Townend DC, Gaudieri S, Dawkins RL (1993) Differences in the central major histocompatibility complex between humans and chimpanzees. Implications for development of autoimmunity and acquired immune deficiency syndrome. Hum Immunol 38:30-41

Lewin HA, Russell GC, Glass EJ (1999) Comparative organization and function of the major histocompatibility complex of domesticated cattle. Immunol Rev 167:145-158

Little CC, Tyzzer EE (1916) Further experimental studies on the inheritance of susceptibility to a transplantable tumor, carcinoma (JWA) of the Japanese waltzing mouse. J. Med. Res. 33:393-453

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155

Makinen A, Chowdhary B, Mahdy E, Andersson L, Gustavsson I (1989) Localization of the equine major histocompatibility complex (ELA) to chromosome 20 by in situ hybridization. Hereditas 110:93-96

Malfroy L, Roth MP, Carrington M, Borot N, Volz A, Ziegler A, Coppin H (1997) Heterogeneity in rates of recombination in the 6-Mb region telomeric to the human major histocompatibility complex. Genomics 43:226-231

Mallya M, Campbell RD, Aguado B (2002) Transcriptional analysis of a novel cluster of LY-6 family members in the human and mouse major histocompatibility complex: five genes with many splice forms. Genomics 80:113-123

Mann AJ, Abraham LJ, Cameron PU, Robinson W, Giphart MJ, Dawkins RL (1993) The caprine MHC contains DYA genes. Immunogenetics 37:292-295

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH (1997) High throughput fingerprint analysis of large-insert clones. Genome Res 7:1072-1084

Marti E, Szalai G, Antczak DF, Bailey E, Gerber H, Lazary S (1996) The equine major histocompatibility complex. In: Schook LB, Lamont SJ (eds) The Major Histocompatibility Complex Region of Domestic Animal Species. CRC Press, Boca Raton, FL, pp 245-267

Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ (2002) The human and mouse replication-dependent histone genes. Genomics 80:487-498

McCann VJ, McCluskey J, Kelly H, Kay PH, Zilko PJ, Christiansen FT, Dawkins RL (1984) Thyrogastric autoimmunity and MHC associated alleles at the C4 locus in patients with type 1 (insulin-dependent) diabetes. Diabetologia 27 Suppl:124-125

McGuire TC, Leib SR, Lonning SM, Zhang W, Byrne KM, Mealey RH (2000) Equine infectious anaemia virus proteins with epitopes most frequently recognized by cytotoxic T lymphocytes from infected horses. J Gen Virol 81:2735-2739

McGuire TC, Zhang W, Hines MT, Henney PJ, Byrne KM (1997) Frequency of memory cytotoxic T lymphocytes to equine infectious anemia virus proteins in blood from carrier horses. Virology 238:85-93

McShane RD, Gallagher DS, Jr., Newkirk H, Taylor JF, Burzlaff JD, Davis SK, Skow LC (2001) Physical localization and order of genes in the class I region of the bovine MHC. Anim Genet 32:235-239

Meredith D, Elser AH, Wolf B, Soma LR, Donawick WJ, Lazary S (1986) Equine leukocyte antigens: relationships with sarcoid tumors and laminitis in two pure breeds. Immunogenetics 23:221-225

MHC Sequencing Consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. Nature 401:921-923

Michalova V, Murray BW, Sultmann H, Klein J (2000) A contig map of the Mhc class I genomic region in the zebrafish reveals ancient synteny. J Immunol 164:5296-5305

Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M (2000) HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proc Natl Acad Sci USA 97:2709-2714

Miller MM, Goto RM, Taylor RL, Jr., Zoorob R, Auffray C, Briles RW, Briles WE, Bloom SE (1996) Assignment of Rfp-Y to the chicken major histocompatibility complex/NOR microchromosome and evidence for high-frequency recombination associated with the nucleolar organizer region. Proc Natl Acad Sci USA 93:3958-3962

Monaco JJ (1992a) Genes in the MHC that may affect antigen processing. Curr Opin Immunol 4:70-73

Monaco JJ (1992b) Major histocompatibility complex-linked transport proteins and antigen processing. Immunol Res 11:125-132

Monaco JJ (1992c) A molecular model of MHC class-I-restricted antigen processing. Immunol Today 13:173-179

Mornet E, Crete P, Kuttenn F, Raux-Demay MC, Boue J, White PC, Boue A (1991) Distribution of deletions and seven point mutations on CYP21B genes in three clinical forms of steroid 21-hydroxylase deficiency. Am J Hum Genet 48:79-88

Muggli-Cockett NE, Stone RT (1988) Identification of genetic variation in the bovine major histocompatibility complex DR beta-like genes using sequenced bovine genomic probes. Anim Genet 19:213-225

Neefjes JJ, Momburg F, Hammerling GJ (1993) Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. Science 261:769-771

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA 94:7799-7806

Nishimura M, Obayashi H, Mizuta I, Hara H, Adachi T, Ohta M, Tegoshi H, Fukui M, Hasegawa G, Shigeta H, Kitagawa Y, Nakano K, Kaji R, Nakamura N (2003) TNF, TNF receptor type 1, and allograft inflammatory factor-1 gene polymorphisms in Japanese patients with type 1 diabetes. Hum Immunol 64:302-309

Nonaka M, Matsuo M, Naruse K, Shima A (2001) Comparative genomics of medaka: the major histocompatibility complex (MHC). Mar Biotechnol (NY) 3:S141-144

Ohta Y, McKinney EC, Criscitiello MF, Flajnik MF (2002) Proteasome, transporter associated with antigen processing, and class I genes in the nurse shark Ginglymostoma cirratum: evidence for a stable class I region and MHC haplotype lineages. J Immunol 168:771-781

Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF (2000) Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proc Natl Acad Sci USA 97:4712-4717

Okamoto K, Makino S, Yoshikawa Y, Takaki A, Nagatsuka Y, Ota M, Tamiya G, Kimura A, Bahram S, Inoko H (2003) Identification of I kappa BL as the second major histocompatibility complex-linked susceptibility locus for rheumatoid arthritis. Am J Hum Genet 72:303-312

Otto SP, Yong P (2002) The evolution of gene duplicates. Adv Genet 46:451-483

Park YH, Joo YS, Park JY, Moon JS, Kim SH, Kwon NH, Ahn JS, Davis WC, Davies CJ (2004) Characterization of lymphocyte subpopulations and major histocompatibility complex haplotypes of mastitis-resistant and susceptible cows. J Vet Sci 5:29-39

Partanen J, Kere J, Wessberg S, Koskimies S (1989) Determination of deletion sizes in the MHC-linked complement C4 and steroid 21-hydroxylase genes by pulsed-field gel electrophoresis. Genomics 5:345-349

Payne R, Rolfs MR (1958) Fetomaternal leukocyte incompatibility. J. Clin. Invest. 37:1756-1763

Peelman LJ, Chardon P, Vaiman M, Mattheeuws M, Van Zeveren A, Van de Weghe A, Bouquet Y, Campbell RD (1996) A detailed physical map of the porcine major histocompatibility complex (MHC) class III region: comparison with human and mouse MHC class III regions. Mamm Genome 7:363-367

Perez J, Mozos E, Martin MP, Day MJ (1999) Immunohistochemical study of the inflammatory infiltrate associated with equine squamous cell carcinoma. J Comp Pathol 121:385-397

Phillips RB, Zimmerman A, Noakes MA, Palti Y, Morasch MR, Eiben L, Ristow SS, Thorgaard GH, Hansen JD (2003) Physical and genetic mapping of the rainbow trout major histocompatibility regions: evidence for duplication of the class I region. Immunogenetics 55:561-569

Raudsepp T, Kijas J, Godard S, Guerin G, Andersson L, Chowdhary BP (1999) Comparison of horse chromosome 3 with donkey and human chromosomes by cross-species painting and heterologous FISH mapping. Mamm Genome 10:277-282

Renard C, Vaiman M, Chiannilkulchai N, Cattolico L, Robert C, Chardon P (2001) Sequence of the pig major histocompatibility region containing the classical class I genes. Immunogenetics 53:490-500

Reyburn H, Mandelboim O, Vales-Gomez M, Sheu EG, Pazmany L, Davis DM, Strominger JL (1997) Human NK cells: their ligands, receptors and functions. Immunol Rev 155:119-125

Romeike A, Brugmann M, Drommer W (1998) Immunohistochemical studies in equine recurrent uveitis (ERU). Vet Pathol 35:515-526

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, NJ, pp 365-386

Sambrook JG, Russell R, Umrania Y, Edwards YJ, Campbell RD, Elgar G, Clark MS (2002) Fugu orthologues of human major histocompatibility complex genes: a genome survey. Immunogenetics 54:367-380

Sato A, Figueroa F, Murray BW, Malaga-Trillo E, Zaleska-Rutczynska Z, Sultmann H, Toyosawa S, Wedekind C, Steck N, Klein J (2000) Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics 51:108-116

Shepherd JC, Schumacher TN, Ashton-Rickardt PG, Imaeda S, Ploegh HL, Janeway CA, Jr., Tonegawa S (1993) TAP1-dependent peptide translocation in vitro is ATP dependent and peptide selective. Cell 74:577-584

Shiina T, Inoko H, Kulski JK (2004a) An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens 64:631-649

Shiina T, Shimizu C, Oka A, Teraoka Y, Imanishi T, Gojobori T, Hanzawa K, Watanabe S, Inoko H (1999a) Gene organization of the quail major histocompatibility complex (MhcCoja) class I gene region. Immunogenetics 49:384-394

Shiina T, Shimizu S, Hosomichi K, Kohara S, Watanabe S, Hanzawa K, Beck S, Kulski JK, Inoko H (2004b) Comparative genomic analysis of two avian (quail and chicken) MHC regions. J Immunol 172:6751-6763

Shiina T, Tamiya G, Oka A, Takishima N, Inoko H (1999b) Genome sequencing analysis of the 1.8 Mb entire human MHC class I region. Immunol Rev 167:193-199

Shiina T, Tamiya G, Oka A, Takishima N, Yamagata T, et al. (1999c) Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region. Proc Natl Acad Sci USA 96:13282-13287

Sinnott P, Collier S, Costigan C, Dyer PA, Harris R, Strachan T (1990) Genesis by meiotic unequal crossover of a de novo deletion that contributes to steroid 21-hydroxylase deficiency. Proc Natl Acad Sci USA 87:2107-2111

Skow LC, Snaples SN, Davis SK, Taylor JF, Huang B, Gallagher DH (1996) Localization of bovine lymphocyte antigen (BoLA) DYA and class I loci to different regions of chromosome 23. Mamm Genome 7:388-389

Smith TP, Rohrer GA, Alexander LJ, Troyer DL, Kirby-Dobbels KR, Janzen MA, Cornwell DL, Louis CF, Schook LB, Beattie CW (1995) Directed integration of the physical and genetic linkage maps of swine chromosome 7 reveals that the SLA spans the centromere. Genome Res 5:259-271

Snell GD (1948) Methods for the study of histocompatibility genes. J. Genet. 49:87-108

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. Comput Appl Biosci 13:523-535

Stephens R, Horton R, Humphray S, Rowen L, Trowsdale J, Beck S (1999) Gene organisation, sequence variation and isochore structure at the centromeric boundary of the human MHC. J Mol Biol 291:789-799

Stone RT, Muggli-Cockett NE (1990) Partial nucleotide sequence of a novel bovine major histocompatibility complex class II beta-chain gene, BoLA-DIB. Anim Genet 21:353-360

Stone RT, Muggli-Cockett NE (1993) BoLA-DIB: species distribution, linkage with DOB, and northern analysis. Anim Genet 24:41-45

Sulston J, Mallett F, Staden R, Durbin R, Horsnell T, Coulson A (1988) Software for genome mapping by fingerprinting techniques. Comput Appl Biosci 4:125-132

Takahashi K, Rooney AP, Nei M (2000) Origins and divergence times of mammalian class II MHC gene clusters. J Hered 91:198-204

Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247-250

Tay GK, Leelayuwat C, Chorney MJ, Cattley SK, Hollingsworth PN, Witt CS, Daly LN, Hughes A, Dawkins RL (1997) The MHC contains multiple genes potentially relevant to hemochromatosis. Immunogenetics 45:336-340

Terado T, Okamura K, Ohta Y, Shin DH, Smith SL, Hashimoto K, Takemoto T, Nonaka MI, Kimura H, Flajnik MF, Nonaka M (2003) Molecular cloning of C4 gene and identification of the class III complement region in the shark MHC. J Immunol 171:2461-2466

Totaro A, Grifa A, Carella M, Rommens JM, Valentino MA, Roetto A, Zelante L, Gasparini P (1998a) Cloning of a new gene (FB19) within HLA class I region. Biochem Biophys Res Commun 250:555-557

Totaro A, Roetto A, Rommens JM, Grifa A, Carella M, d'Agruma L, Valentino MA, D'Ambrosio L, Cicilano M, Camaschella C, Franco B, Gasparini P (1998b) Generation of a transcription map of a 1 Mbase region containing the HFE gene (6p22). Eur J Hum Genet 6:105-113

Trowsdale J (1995) "Both man & bird & beast": comparative organization of MHC genes. Immunogenetics 41:1-17

Trowsdale J (2001) Genetic and functional relationships between MHC and NK receptor genes. Immunity 15:363-374

Tusie-Luna MT, White PC (1995) Gene conversions and unequal crossovers between CYP21 (steroid 21-hydroxylase gene) and CYP21P involve different mechanisms. Proc Natl Acad Sci USA 92:10796-10800

van der Poel JJ, Groenen MA, Dijkhof RJ, Ruyter D, Giphart MJ (1990) The nucleotide sequence of the bovine MHC class II alpha genes: DRA, DOA, and DYA. Immunogenetics 31:29-36

van Rood JJ, Eernisse JG, van Leeuwen A (1958) Leukocyte antibodies in sera of pregnant women. Nature 181:1735-1736

Velten FW, Renard C, Rogel-Gaillard C, Vaiman M, Chardon P (1999) Spatial arrangement of pig MHC class I sequences. Immunogenetics 49:919-930

Wagner JL (2003) Molecular organization of the canine major histocompatibility complex. J Hered 94:23-26

Wagner JL, Burnett RC, Storb R (1999) Organization of the canine major histocompatibility complex: current perspectives. J Hered 90:35-38

Walter L, Gunther E (2000) Physical mapping and evolution of the centromeric class I gene-containing region of the rat MHC. Immunogenetics 51:829-837

Walter L, Hurt P, Himmelbauer H, Sudbrak R, Gunther E (2002) Physical mapping of the major histocompatibility complex class II and class III regions of the rat. Immunogenetics 54:268-275

Watts C (2001) Antigen processing in the endocytic compartment. Curr Opin Immunol 13:26-31

Wright H, Ballingall KT (1994) Mapping and characterization of the DQ subregion of the ovine MHC. Anim Genet 25:243-249

Wright H, Ballingall KT, Redmond J (1994) The DY sub-region of the sheep MHC contains an A/B gene pair. Immunogenetics 40:230-234

Xie T, Rowen L, Aguado B, Ahearn ME, Madan A, Qin S, Campbell RD, Hood L (2003) Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse. Genome Res 13:2621-2636

Xu A, van Eijk MJ, Park C, Lewin HA (1993) Polymorphism in BoLA-DRB3 exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus. J Immunol 151:6977-6985

Yang Z, Yu CY (2000) Organizations and gene duplications of the human and mouse MHC complement gene clusters. Exp Clin Immunogenet 17:1-17

Yeager M, Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol Rev 167:45-58

Yoshino M, Xiao H, Jones EP, Kumanovics A, Amadou C, Fischer Lindahl K (1997) Genomic evolution of the distal Mhc class I region on mouse Chr 17. Hereditas 127:141-148

Yu CY, Whitacre CC (2004) Sex, MHC and complement C4 in autoimmune diseases. Trends Immunol 25:694-699

Yuhki N, Beck T, Stephens RM, Nishigaki Y, Newmann K, O'Brien SJ (2003) Comparative genome organization of human, murine, and feline MHC class II region. Genome Res 13:1169-1179

Yung Yu C, Yang Z, Blanchong CA, Miller W (2000) The human and mouse MHC class III region: a parade of 21 genes at the centromeric segment. Immunol Today 21:320-328

Zemmour J, Koller BH, Ennis PD, Geraghty DE, Lawlor DA, Orr HT, Parham P (1990) HLA-AR, an inactivated antigen-presenting locus related to HLA-A. Implications for the evolution of the MHC. J Immunol 144:3619-3629

APPENDIX A

Table A-1 BAC clones identified in the initial screening of the CHORI-241 library.

CHORI-241 BAC CLONE ADDRESS	GENE CONTENT
004 H14	HSP70
005 G21	EC1
006 E13	PBX2
008 K14	DRB, DQB, DOB
012 F15	DRB, DQB, DRA
012 L22	DRB, DQB, DQA,DRA
013 C16	DRB, DQB, DRA
014 D12	EC1
021 M20	EC1
022 H19	POU5F1, TCF19
022 H20	DRB, DQB, DQA,DRA
023 F07	HSP70
024 A13	POU5F1, TCF19, EC1
25 L07	BAT3, TNF
030 B24	EC1
032 O03	DRB, DQB, DQA,DRA
040 F20	EC1
040 O14	EC1
042 P05	----------
061 B19	MOG
63 F17	GNL1
066 F10	HSP70
069 N18	EC1
073 L13	DRB, DQB, DQA,DRA
074 D13	EC1, EC2
097 H06	POU5F1, TCF19
102 G05	EC1, RNF9, ZNF173
105 A11	GNL1
111 P20	----------
120 O23	POU5F1, TCF19
120 J04	EC1
132 I24	DRB, DQB, DOB
135 M23	DOB, DRB, DQA, DQB
136 H02	DRB, DQB, DOB
138 C23	MOG
147 K21	BRD2, PSMB8, DMB
150 F11	EC1
158 E20	DRB, DQB, DRA
163 J11	RXRB
163 M01	DRB, DQB, DOB
163 G18	HSP70
164 I22	PBX2
167 M09	EC1, EC2
168 B14	DRB, DQB, DOB

Table A-1 Continued

CHORI-241 BAC CLONE ADDRESS	GENE CONTENT
168 K09	HSP70
169 B13	DRB, DQB, DQA
169 I13	HSP70
172 N21	HSP70
174 H21	EC1
186 O13	HSP70
189 L11	DRB, DQB, DOB
190 H04	EC1
288 J19	DRA, DRB, DQA, DQB
324 F02	Class I
337 M04	BAT3, TNF
341 J05	PBX2
345 J21	DRB, DQB, DRA
346 D19	HSP70
346 D20	HSP70
348 L13	COL11A2
348 K19	PBX2
352 NO5	DRB, DQB
355 H03	EC1, EC2
356 H20	FLOT1
359 L18	BRD2, PSMB8, DMB
360 H03	DRB, DQB, DQA,DRA
363 F17	POU5F1, TCF19
363 O17	DRB, DQB, DRA
371 L22	HSP70
371 F04	MOG
372 F10	PBX2
373 P04	DRB, DQB, DRA
382 H22	MOG
384 D16	HSP70
389 G09	DRB, DQB
394 D17	BAT3, TNF
400 G19	DRB, DQB, DOB
406 C23	GNL1
407 K07	BTNL2
408 P17	BRD2, PSMB8, DMB
410 K24	RXRB, BRD2
412 N14	GNL1, EC1
416 B09	GNL1
418 J06	GNL1
418 N06	GNL1
426 M04	EC1
431 P04	HSP70
439 J08	BRD2, PSMB8, DMB
439 E06	PPP1R10, ABCF1, GNL1
440 J07	DQA, DQB
441 N13	DRB, DQB, DQA

Table A-1 Continued

CHORI-241 BAC CLONE ADDRESS	GENE CONTENT
454 N08	DRB, DQB, DRA
455 C07	DOB, DRB
464 F03	POU5F1, TCF19
464 C16	POU5F1, TCF19
475 L06	FLOT1
486 D08	POU5F1, TCF19
492 I22	COL11A2
497 G11	HSP70
498 L17	DRB, DQB, DRA
499 G05	GNL1
499 L11	FLOT1
510 G12	DRB, DQB, DQA
510 E03	HSP70
516 O02	GNL1
516 J05	MOG
522 P24	PBX2
523 H09	DRB, DQB, DQA,DRA
528 G20	DRB, DQB, DOB
528 E24	MOG

All gene assignments were confirmed by Southern blot hybridization. Genes listed in red were added based upon Southern blot hybridization results. Genes listed in blue indicate questionable gene assignment by Southern blot hybridization results. BAC clones addresses listed in bold were identified and confirmed by collaborators at Cornell University.

VITA

Ashley Gustafson Seabury

1801-B Woodsman Dr.

College Station, TX 77840

Education

2005 Ph.D., Genetics

Texas A&M University, Department of Veterinary Integrative Bioscience

2001 B.S., Genetics

West Virginia University, College of Agriculture and Consumer Science

Summa Cum Laude

Publications

Rexroad CE 3rd, Coleman RL, Gustafson AL, Hershberger WK, Killefer J (2002) Development of rainbow trout microsatellite markers from repeat enriched libraries. Mar Biotechnol 4:12-16

Gustafson AL, Tallmadge RL, Ramlachan N, Miller D, Bird H, Antczak DF, Raudsepp T, Chowdhary BP, Skow LC (2003) An ordered BAC contig map of the equine major histocompatibility complex. Cytogenet Genome Res 102:189-195

Gustafson-Seabury A, Raudsepp T, Goh G, Kata SR, Wagner ML, Tozaki T, Mickelson JR, Womack JE, Skow LC, Chowdhary BP (2005) High-resolution RH of horse chromosome 22 reveals a putative ancestral vertebrate chromosome. Genomics 82:188-200